Humans in the Loop:

The evolution of work in early experiments with Generative AI

April 2026

Summary findings from the
MIT Working Group on Generative AI and the Work of the Future

MIT industrial PERFORMANCE CENTER

Authors
Ben Armstrong
Julie Shah

Contributors
Kate Kellogg
Sabiyyah Ali
Greer Brigham
Carey Goldberg
Shakked Noy
Prerna Ravi
Azfar Sulaiman
Leana Tejedor
Felix Wang
Whitney Zhang

EXECUTIVE SUMMARY

How can generative AI lead to better jobs? In January 2026, approximately half of American workers reported using AI, but how new technologies have affected the quality of their jobs remains largely unclear. Drawing on a study of more than twenty companies across four major industry groups – healthcare and life sciences; retail; finance, insurance, and real estate; and manufacturing – this paper identifies patterns in how organizations have experimented with generative AI; how those experiments have changed the roles of workers; and how organizations can support high-quality jobs as they integrate new technologies.

The applications of generative AI among the companies we studied were directed toward three common challenges. The *bottleneck problem* is where workers are responsible for a growing number of simple tasks that get in the way of higher value-added work. Generative AI tools have been aimed at relieving these bottlenecks by speeding up near-routine tasks. The *cafeteria problem* emerges in a process that requires workers to consult experts from various domains and integrate their input into a product, document, or idea. Organizations have looked to generative AI to predict what those domain experts might have said based on what they have produced in the past. The *learning curve* problem represents the extra time and effort novices require to complete a complex task in a new domain. Generative AI tools have been directed to help workers to perform as if they had more experience – and to develop expertise – in new domains.

Across the applications of generative AI addressing these challenges, there has been a shift in the core tasks that professional and technical workers are being asked to perform. Where generative AI tools are being deployed, workers are increasingly asked to perform supervisory control tasks as the "human in the loop" overseeing and analyzing a process rather than executing the process manually. Although supervisory control tasks may be new to workers in law or healthcare, the "human in the loop" concept is not new. A range of occupations from airline pilots and manufacturing technicians to utility operators are supervisors of automated systems, and there are guidelines for how workers in these roles can thrive that can inform the use of generative AI.

Supervisory control jobs vary widely in their quality and compensation. Whereas operators of complex systems in nuclear power and aerospace are widely considered interesting and well-paid (if intense) roles, machine operators overseeing automated equipment in industrial environments frequently receive lower pay and are harder for employers to fill. There is a similar range of emerging jobs in generative AI environments: some generative AI tools may require humans in the loop to perform tedious tasks reviewing content that generative AI has produced, whereas other roles call for supervisory control work to interpret and troubleshoot information, which requires higher skill and more engagement.

Employees have significant discretion over how they use generative AI tools – and can often shape how they affect their daily work. Although public discussions have frequently presented the march of generative AI as an inevitable force to which companies must react, there has been significant variation in how organizations have chosen to deploy generative AI within their organizations. This variation affirms that management practices matter for how generative AI will shape jobs.

Employers, policymakers, and technology providers have several levers to make jobs affected by generative AI more interesting and higher quality: designing interfaces for transparency and situational awareness, designing jobs for learning and mobility, and supporting training that develops judgment and domain expertise. On each of these points, public policy can make it easier for organizations to pursue these practices.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The MIT working group on Generative AI and the Work of the Future was made possible with grants from Google's Digital Future Initiative and the Siegel Family Endowment, as well as the generous participation of more than 50 organizations in research interviews, working group meetings, and events. A special thanks goes to those who sat through interviews, described their work, and reflected on their experiments with generative AI. The working group also benefited from the hard work and creative energy of terrific graduate students and research fellows listed as contributors to this study.

1. INTRODUCTION

Soon after ChatGPT was released to the public in November 2022, Mark started peppering it with questions.

“What are you good at?”

“What makes you happy?”

“How do you have world peace with finite resources?

“Are you alive?”

Mark has a playful attitude toward new gadgets. An affable Midwesterner, he wasn’t afraid of ChatGPT as much as he was curious about how it actually worked – and what he could do with it.

As an engineering director at a Fortune 500 company, Mark was already familiar with AI. He had been working on integrating machine learning algorithms into his employer’s business for years. He had seen incremental progress, but nothing transformative.

Everyone was telling him that this time was different. Generative AI was going to change everything.

Part of him was skeptical of the hype. He had been burned before. The previous year, everyone in his field was saying that the metaverse was coming. He had worked on preparing his company to use VR headsets. That was a bust.

But at the same time, he was reading respected technology voices talking about Generative AI as a big step toward an Artificial General Intelligence where the algorithm could reason on its own, unprompted. He hadn’t heard people get this excited since the dawn of the internet. Mark wanted to see what was there -- how intelligent was this thing, really?

The answers the algorithm gave were nicely written and organized, but they weren’t that interesting. The way that ChatGPT described its skills was generic. Its response to his question about resource scarcity was word salad. It used human-like language, but it wasn’t anything like a human.

He kept poking and prodding, following up on the answers ChatGPT gave him. It became clear the bot wasn’t thinking on its own. He didn’t have to worry about anything like consciousness. But it could be an incredibly effective assistant.

He started asking the algorithm to rewrite emails and memos, to summarize complex documents for him. Then he started using it as he developed new software programs – to perform unit tests (a necessary task, but one he dreaded) and to help him develop a front-end for an app he was developing (an area that wasn’t his specialty). Each time he hit a bump or didn’t know the answer, he could ask the bot. It might not give him everything he needed, but it would get him moving again. Before, he might have gone down a rabbit hole, calling a friend or scrolling an online forum. Now, it was faster, and he felt he was capable of doing more.

What began as a curiosity became a legitimate hobby. Outside of work, Mark was watching YouTube videos and reading blogs about how best to prompt ChatGPT. He started teaching his team at work what it could do for them. It wasn’t just that the tool could help remove the drudgery – the repetitive coding – from their daily work. They could also develop new technology that reached customers.

Around the same time Mark and millions of others began experimenting with ChatGPT, the heads of companies large and small were imagining how the technology might change their businesses. They knew that rank-and-file workers, particularly engineers, were already playing with the technology and seeing how it could help them. They wanted to harness that creative spirit, but they also worried.

No one quite knew how the models worked, or why they gave the responses they did. You could prompt ChatGPT with the same question two times in a row, and it would give you different responses. Corporate leaders had heard that the technology could “hallucinate,” offering responses that were unreliable. They were also worried that if employees put confidential data into the model, it could show up in a response to a competitor’s query.

Many organizations put up walls, at least for a little while. Legal departments sent company-wide blasts with regulations asking companies not to use ChatGPT on the company network. The firewalls went up, and companies formed task forces on how they could use the tool responsibly.

Three years after Mark began experimenting, organizations and individuals are still proving the concept of AI at work – laboring to transform processes and jobs beyond pilot demonstrations. It's unclear for how long the ground is still going to be shifting beneath them. Bill Gates wrote in 1996, as Microsoft was taking off, "We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten."

The generative AI transformations that companies have embarked on over the past three years are part of a decades-long journey. This paper is a chronicle of how that journey has begun – the experiments, false starts, and visions of those first years, as seen through the eyes of companies on the frontier of that journey.

The value of these stories in progress is to capture key decisions and motivations of the leaders in the field as they set out on their long-term paths. In some cases, the lessons from their experiences might be motivation to course correct, or re-evaluate assumptions that have become taken for granted along the way. In other cases, this time capsule may serve as a reminder of the goals and tradeoffs that companies identify as they embark on their varied paths.

It has become a common assumption among AI startups, the press, and the public that Generative AI represents a revolutionary disruption to the labor market and the economy. Since AI tools can perform a growing body of tasks as well as, or better than, human workers, there have even been projections that artificial general intelligence (AGI), where computers can do just about any job as well as a human, is just around the corner.

Business leaders and policymakers frame generative AI tools as an exogenous shock – the type of dramatic event, almost like a hurricane – that organizations can prepare for and adapt to, but one where managers, workers, and leaders do not have substantial discretion in shaping. It has become common among business leaders to paraphrase Nvidia CEO Jensen Huang, who predicted, "AI will not take your job, but someone who knows AI will."

Economic analysts have projected that generative AI threatens white collar jobs, often occupied by skilled workers, and that middle-skill jobs in the skilled trades may become beneficiaries of this wave of technological change. The idea that generative AI might threaten skilled work draws on the assumption that if generative AI tools can perform a skilled task for which a professional is compensated, that professional's job is under threat. This was the assumption that led AI pioneer Geoffrey Hinton to warn that medical schools should stop training radiologists.

Although it is too early to evaluate any of these predictions fully, there are significant stakes in each of them. Rapid and significant displacement of workers would require a much different policy and educational response compared to technological changes of the past. If business leaders and workers had less discretion and insight into how their technology tools worked, then there would need to be a new set of principles and methods for integrating new technologies and evaluating their usefulness. And if the effects of new technologies were suddenly skill leveling rather than skill biased, there would be significantly different incentives in the labor market that might prompt employers to reorganize their workforce dramatically.

This study is focused on a question that underpins each of the common predictions about generative AI – and that we can begin to answer with data currently available: how can generative AI technologies support better jobs? This has long been the interest of research focused on technology and work. There is a search for the levers that policymakers, workers, and employers can pull to harness the possibilities of new technologies to create benefits shared by workers and firms alike. What "better jobs" look like is not always a point of consensus, but in this paper we consider improvements in wages, upward mobility, and safety, as well as jobs that are more engaging and more flexible, to be better.

To understand how the early uses of generative AI tools were changing jobs, our research examined how a group of large employers, primarily in healthcare, retail, manufacturing, and FIRE (finance, insurance, real estate), experimented with these technologies. Through interviews with technical leaders developing AI applications, as well as managers and workers on teams using generative AI tools, we observed the variety of approaches that firms – and workers within them – took to making these new technology tools work within their organizations.

Three questions were of particular importance to us. What tasks were organizations asking generative AI to do – and how were workers' roles changing as a result? Which types of generative AI applications were scaling up, and which types were losing steam? And what new skills and training was the introduction of AI tools requiring of workers?

We found common threads in the applications of generative AI across industries. For example, the uses of generative AI for marketing, customer service, and legal research apply in finance, as well as in retail. The employers we studied frequently used generative AI to solve some combination of three problems, which we defined as accelerating annoying, near-routine tasks (the bottleneck problem), integrating information from diverse experts (the cafeteria problem) and achieving fluency quickly in a new domain (the learning curve problem). The applications of generative AI to address these problems have manifested differently across industries, but the problems and the general shape of the solutions appear similar.

Across applications of generative AI, the role for the human worker shifted away from executing a task and toward supervising a task. This transition is consistent with patterns of automation and its impact on work since the deployment of early autopilot systems in the 1950s. Just as a machinist transitioned from manually operating a mill to overseeing a mill executing a computer program with the introduction of Computer Numerically Controlled (CNC) machining, customer service representatives are in some cases shifting from having conversations on their own with customers to overseeing a customer's interaction with a bot.

While the capabilities of generative AI to augment certain tasks is certainly new, the principles of supervisory control applied to automated systems are well developed and can provide a guide for the successful use of AI. Decades of scholarship on human-machine interaction in the digital age emphasizes the role of situational awareness, trust, and mental workload in supporting high performance among human workers. These factors can also be important to job quality. Too much workload or not enough trust could drive workers to leave an otherwise high-quality job.

Although we only observed the early stages of the scale-up of generative AI tools, there was a consistent difference between mature AI projects and pilots. The most mature AI projects that companies shared with us had a combination of three characteristics: they addressed a business problem that the organization had been trying to solve *before* the introduction of generative AI tools. They required a combination of technologies – not just generative AI – to perform the relevant task. And these mature applications often required buy-in and feedback from domain experts close to the process at hand.

We did not find consistent evidence on the impact of generative AI tools on skill requirements, nor did we find a consensus on the training required. Some applications of generative AI were aimed at reducing the skills required to perform a task – including in cases where generative AI might help people climb the learning curve more quickly – whereas other tasks integrating generative AI required as much or more domain expertise from the humans in the loop. The evidence suggests that there is substantial discretion among employers – and even the designers of generative AI tools – over how the skill demands within their organizations evolve.

Returning to the core question – how can AI support better jobs? – the answer is up to employers, which decide how to use generative AI tools – and shape the impact that those tools have on jobs. Employers have an interest in discovering how to apply generative AI in a way that improves the performance of their businesses, while also increasing flexibility and opportunity for workers – a balance that we call positive-sum automation.

The levers for employers are in the interfaces of generative AI tools, which can be designed so that workers are prepared to learn and improve in their jobs. Human factors research shows how interfaces designed to enhance workers' awareness and balance their workforce can support high performance in high-stakes settings like aviation. Employers can also deploy generative AI tools in ways that are transparent and build trust with workers, which make workers more likely to champion the uses of new technologies that can improve the organization's performance. And finally, the introduction of new technologies consistently presents an opportunity to reconfigure processes and create new types of work. It is not necessarily up to AI startups to create the new jobs that emerge from the influx of new technology. It can be up to users of these technologies to understand the types of jobs that will position humans in the loop to perform at their best.

This paper summarizes our research on large companies' experimentation with generative AI in three parts. The first lays out the hypotheses that we were aiming to test at the outset. There was speculation that "this time is different" when it comes to the impact of new technologies on work. We compare contemporary analyses to historical understandings of how technologies affect the labor market. The second section reviews what we found from extensive interviews and working group meetings with companies experimenting with generative AI. We break our findings down by industry, as well as by the research question of interest. The third and final section draws lessons from the findings for practitioners. The focus is on how historical experience developing automated systems – combined with recent experience with generative AI – might combine to offer a framework for how companies can create new work that represents positive-sum automation.

2. HYPOTHESES: HOW THIS TIME [MIGHT BE] DIFFERENT

At a hearing under the stately arches of the Old Supreme Court chamber in the U.S. Capitol Building, lawmakers, business leaders, and scholars gathered for a special hearing on a technological breakthrough. Witness after witness described how new computing technologies promised to extend human capabilities and automate manual work. The upside was huge, they argued. Workers could produce far more per hour, and the technologies could help give rise to entirely new industries. One of the main questions was whether the United States could compete with its global foes and win the race to control the trajectory of the technology into the future.

While optimism filled the air, some urged caution. "If it were not for our present high-level employment, prosperous economic situation," one Senator said, there might be "painful adjustments as a result of the great rush of technological change." A statement from a labor leader emphasized that the change would require real adjustments from workers: what new skills would be required? What retraining would be offered? Would there be "large-scale social dislocations?"

"Questions of designing, installing and maintaining" are all problems that can be "solved by the engineer," the labor leader's statement read. "It is the human problems, the questions of adjustment to change, or preparation for new assignment…that cannot be solved in such a mechanical fashion but must be subject to the most imaginative thinking that all of us can give."

The year was 1956. The labor leader was George Meany, then-President of the recently formed AFL-CIO.

Over the last 70 years, the public debate over new technologies and their impact on the economy have followed a familiar script. There is excitement about the prospects for economic growth and productivity on the one hand; concerns about what the change will mean for workers and incumbent industries on the other. An abiding question has been: how can new technologies support higher performance for employers, while also enabling higher-quality jobs for workers?

Wave after wave of technological change have produced consistent lessons. First, the impact of new technologies on jobs is often evolutionary – not revolutionary. Although technological breakthroughs can seem to introduce rapid change, the impact of those innovations often filters into businesses and the labor market slower as employers take time to transform their processes, redesign their jobs, and identify the most promising ways to use technology tools. Contrary to frequent predictions about massive job displacement due to automation technologies, technological change is often associated with incremental increases in some jobs and decreases in other jobs. At the margins, new jobs emerge and other jobs disappear, but there is frequently time for trainees and incumbent workers to identify new careers. This is in contrast to other economic shocks, like recessions or plant closures due to offshoring, which can dislocate large groups of workers at once.

Second, management practices and design decisions can shape how new technologies affect workers. The introduction of new machines and software have not had a uniform effect across the organizations that adopt them. In some organizations, new technologies are harnessed as an opportunity to upskill workers and create higher-wage jobs. In other organizations, technologies might present a chance to deskill or commodify work, adapting to a workforce with higher turnover or less training. In some cases, the decision of how to adopt technology has been reserved for managers planning how to redesign work and integrate new technology tools. In other cases, the design of the technology tools themselves have affected how they affect workers. Some technologies, for example, have been designed as closed systems so that workers using them have little discretion over how to use them or where they can be improved. In other cases, more flexible technology design allows workers to innovate and tinker on their own have allowed for bottom-up innovation and improvements.

Third, education and training often influence how technological change affects the workforce. The "race between technology and education" indicates that when the educational attainment of the workforce progresses slower than technology, the most educated will reap disproportionate advantages from technological change. But when educational attainment grows more quickly, the fruits of technological change can be shared more evenly. For evidence, consider the 1950s and 1960s, when the rise of early computing technology coincided with a growing population of high school and college graduates, as well as the golden era of the American middle class.

Beginning in the 1980s, the expansion of personal computers was associated with a different sort of transformation – a skill-biased technological change that seemed to advantage the most educated and limit job prospects for the rest of the labor market. Since 2010, as enrollment in higher education has declined, there has been concern that new computing technologies and advances in automation hardware like robotics may exacerbate the polarization of the labor market that has seen high wage growth for educated workers and the hollowing out of middle-wage job opportunities.

Since late 2022, as business leaders, policymakers and scholars have experimented with and gathered data on the impact of generative AI tools, there has been justifiable awe at the capabilities of generative AI tools, as well as the speed with which they have improved. There has also been a frenzy of predictions that this moment of technological change is somehow different – and that the lessons of past technological changes may not apply to this one.

Artificial Intelligence is "more profound than fire, electricity, or anything that we have done in the past," Google CEO Sundar Pichai says. Bill Gates calls it "the most transformative technology any of us will see in our lifetimes." And Jamie Dimon of JP Morgan Chase expects AI to be at least as world-changing as the steam engine, computing and the Internet.

Darker predictions abound as well: the new, generative AI could lead to mass unemployment and economic dislocation. Prominent researcher Eliezer Yudkowsky has argued that the only way to avoid AI-induced disaster is simply to "shut it all down."

Historic as this moment is, the exclamatory headlines mask a more complicated story. Although the new technology is undeniably wondrous, and the most apocalyptic or rapturous predictions may yet come true, this infancy period of generative AI is not living up to the early proclamations of "this changes everything."

"This is not as easy as it looks" is a common refrain from companies on the uphill climb to adjust to generative AI. Studies are finding a "jagged frontier": AI is far more useful on some tasks than others. And even though the defining feature of generative AI is that it seems to be a general-purpose technology, in some settings it remains a hammer in search of a nail. For many companies, it simply does not fit very well — at least not yet — into the main tasks that sustain their comparative advantage.

Since companies began experimenting with the first wave of generative AI tools in late 2022 and early 2023, there have been two competing sets of hypotheses about the impact that these tools would have on jobs.

The first set of hypotheses predicted that these technologies would behave differently than the technologies of the past. In some ways the "different" hypotheses were familiar, since past technologies like the internet and robotics had similarly inspired predictions of dramatic short-term impact on the labor market, only to generate more incremental changes.

The second set of hypotheses represent expectations based on historical experiences. Consider if the evolution of generative AI and its impact on work unfolded along a parallel path to a platform technology like the personal computer, or an automation technology like an industrial robot. This section introduces three hypotheses predicting how generative AI tools are different from the technologies of the past, and compares them to the hypotheses based on historical evidence.

HYPOTHESIS 1: GENERATIVE AI WILL HAVE A MORE RAPID AND TRANSFORMATIVE IMPACT ON THE WORKFORCE THAN PAST TECHNOLOGIES.

Since generative AI tools are currently capable of performing a wide variety of business tasks – and these tools are improving rapidly – the prediction is that widespread adoption of generative AI is likely to generate high productivity growth and significant displacement of existing jobs. Scholars, business leaders, and policymakers – advocates for and opponent of AI technologies – have offered numerous versions of this hypothesis, ranging from nuanced assessments of *potential* AI disruptions to more ambitious predictions of the labor market impact. Academic papers and consulting firms have estimated high anticipated productivity gains, along with a high share of jobs – ranging from nearly 50% to 70% -- with a substantial share of tasks vulnerable to automation from generative AI and related technologies. AI executives have been more explicit with Anthropic CEO Dario Amodei predicting that AI could "disrupt" 50% of white collar jobs in one to five years, which would be associated with unemployment rates of 10% to 20%. Senator Bernie Sanders, drawing on analysis conducted by generative AI, suggested that generative AI could displace 100 million U.S. jobs by 2035.

While some versions of this hypothesis are sensationalized, the common thread is that a range of voices suggest that the speed and scale of the generative AI impact on the labor market will represent a sharp disruption, rather than incremental change. An early lab experiment on using AI for writing tasks, for example, suggested that individuals using AI could not only do the task faster, but they were also judged to perform the task at a higher quality. These findings have been replicated across fields, showing that workers using AI -- for certain discrete tasks -- can perform the tasks faster and at a higher quality on average than workers not using AI.

However, there has also been competing evidence, which is consistent with the hypothesis that the impact of generative AI on the labor market will be far more incremental, similar to technologies of the past. Part of the evidence for incrementalism relies on the false predictions of the past. For one, Geoffrey Hinton's 2016 prediction that "people should stop training radiologists now" since it was "completely obvious" that AI tools would be superior to human expertise. AI technologies for radiology have improved, and the number of radiologists has grown over this period. Similarly, past economic forecasts of the impact of new technologies, including AI on the labor market, have included similar shares of jobs vulnerable to automation. In a widely-cited 2013 paper, the prediction was that 47% of the labor market in the United States was vulnerable to automation by 2030, only for a follow-up study to conclude that a small change to the forecasting methodology yielded just 9% of jobs similarly vulnerable. Forecasts of the impact of generative AI have similarly varied, including predictions that these tools will only have a modest impact on productivity.

Evidence from past technologies, such as industrial robots and robotic process automation, suggest that just because a technology has the potential to affect jobs, it does not guarantee that the impact will be realized. For instance, industrial robots have the capability of performing many routine physical tasks for manufacturing operators and warehouse workers in the United States. Nonetheless, there has been growth in these job categories even as employers have adopted robots. There are several reasons. The first is that new technologies have filtered into enterprises gradually as they have learned how to make use of them. The integration of new technologies to automate tasks that had previously been managed by humans often requires complex re-organization of business processes and work organization.

The second is that the adoption of new technologies often goes hand in hand with opportunities for growth. There are several studies showing that the firms adopting robots and other automation equipment have on average experienced job growth and profitability growth. This does not mean that investments in automation led to growth, nor does it mean that robot adoption is associated with job growth overall (firms not adopting robots may face decline due to the increased competitiveness of robot adopters). However, the association between investments in automation and job growth within individual firms suggests that firms have not historically used investments in automation to conduct layoffs.

Adjudicating between these competing hypotheses has high stakes for the labor market and the economy at large. However, it is too soon to resolve the extent of the impact of generative AI on the labor market just three years into the adoption of these tools. Moreover, the evidence gathered for this study is focused on the micro-level processes of AI adoption and workforce impact – not their aggregate effects. The contribution that this research can make, however, is to compare the speed and scale of adoption and deployment of these tools to the speed and adoption of previous technologies, such as robots and RPA, to assess just *how different* they seem based on company-level evidence.

HYPOTHESIS 2: THE ADOPTION OF GENERATIVE AI WILL BE SKILL LEVELING, YIELDING MORE BENEFITS TO WORKERS WITH LESS FORMAL EDUCATION.

Whereas past digital technologies have been skill-biased, complementing workers with high levels of education, there have been predictions that generative AI technologies will be skill leveling: using generative AI will provide higher productivity and performance gains to workers with less skill and experience compared to workers with high levels of skill and experience.

The concept behind this hypothesis is simple: generative AI tools allow individuals to access support to perform various tasks, from writing to computer programming to troubleshooting and problem solving. For individuals who are already skilled at a task, the support of the generative AI tool may help them improve performance and speed, but marginal performance improvement is harder to achieve since the starting point is already high. For individuals who are not skilled in a task, generative AI can boost performance through two channels: first, for tasks that require it can help less-skilled individuals learn relevant information to complete the task (e.g. using a generative AI chatbot to troubleshoot a customer problem). Second, it can perform tasks beyond the human's capabilities on their behalf (e.g. using a generative AI coding tool to generate a computer program despite not knowing the syntax of the coding language). In both cases, the tool would appear to be less beneficial to highly-skilled individuals in each domain – who understand how to solve the problem and know how to program in the language – than it would be to relative novices.

There is early evidence from multiple domains to support the skill leveling hypothesis. For writing, consulting, and customer service tasks, individuals with lower skills or less experience saw greater performance gains from using the AI tools than people that already had demonstrated high skill at the task. Even in the classroom, the impact of AI on test scores for prospective law students reflected a similar pattern: law students with the lowest expected scores saw the greatest benefits from AI.

The skill leveling hypothesis has implications for wage equality. As companies adopt generative AI, tasks assigned to different jobs could be reshuffled so that there is a reduction in skill requirements for high-wage jobs. One proposed example of skill leveling is in healthcare, where there is an established skill hierarchy between nurses, nurse practitioners, and medical doctors. If AI tools can enable nurse practitioners to access additional knowledge to extend their abilities to provide patient care, they may be able to take on additional tasks that were previously reserved for a doctor. The nurse practitioner could then earn higher wages for their work based on their new responsibilities. In the aggregate, reshuffling like this could translate into higher demand for middle-wage jobs and increase wage equality in the labor market.

The competing hypothesis is that generative AI tools remain skill biased, similar to digital technologies since the 1980s, which have been associated with wage and job growth for workers with a college degree. Even though it appears there are more significant performance gains for lower-skilled individuals on certain tasks, the adoption of AI tools may still lead to growing demand for highly-skilled workers. For instance, the tasks that generative AI can perform – and the level at which they can perform those tasks – may compete with the work of less-experienced and lower-skilled workers in a particular field. And managing generative AI tools may require higher skilled, more experienced workers.

HYPOTHESIS 3: EXPERTISE WITH GENERATIVE AI IS MORE CRITICAL THAN DOMAIN EXPERTISE AND CAN HELP INDIVIDUALS AND ORGANIZATIONS LEAPFROG EXPERIENCED INCUMBENT.

Nvidia CEO Jensen Huang has popularized this sentiment with the idea, paraphrased, that AI will not take your job, but someone using AI will. There is a corollary for businesses, which face prospective challenges from "AI native" startups that aim to displace established market leaders.

The hypothesis inherent in these concepts is that knowing how to use generative AI effectively can be more powerful than industry experience. The design of early generative AI tools like ChatGPT, Claude, and Gemini emphasizes the range of tasks that generative AI can help perform, as well as the low barriers to entry for making use of these tools. Individuals and organizations can merely prompt a generative AI tool using natural language to perform complex tasks across a variety of domains. Learning how to perform a complex writing, computer programming, or data science task would have traditionally required years of investment in expertise and infrastructure development (e.g. databases, servers), whereas performing these tasks with generative AI could obviate the need to make these investments.

The implication of this hypothesis is that organizations can use generative AI to perform complex tasks like computer programming and customer service without relying on underlying technologies, databases, or expertise to do so. Whereas past technologies have been sold as tools that can build on a technology stack to augment or automate individual tasks – or small groups of tasks – early generative AI companies have presented their products as "copilots" or "agents" that can take on entire workflows with limited structured input or infrastructure provided by the user.

The nearest analogy is the way that communities without landlines leapfrogged that technology straight to mobile phones, and communities without a reliable electrical grid leapfrogged straight to solar power. With generative AI, an organization without a well-developed data science infrastructure may use generative AI tools to summarize and query its unstructured information. For organizations that had well-established internal databases or knowledge graphs, the prediction is that generative AI tools might not just supplant them, but render them useless.

There are ways to test this hypothesis at the individual level and at the enterprise level. At the individual level, a decreasing reliance on domain expertise in the tasks that workers are asked to perform would be early supporting evidence for this hypothesis. At the enterprise level, the relative success of startups with deep technological capabilities, but without significant industry experience, at winning market share from incumbents would support this idea. There are ways to test this hypothesis at the individual level and at the enterprise level. At the individual level, a decreasing reliance on domain expertise in the tasks that workers are asked to perform would be early supporting evidence for this hypothesis. At the enterprise level, the relative success of startups with deep technological capabilities, but without significant industry experience, at winning market share from incumbents would support this idea.

3. GENERATIVE AI MEETS REALITY

In 2023, the MIT working group on generative AI and the work of the future began convening a group of companies to understand how they were approaching the adoption of generative AI tools. We originally targeted a small sample of companies to study their uses of generative AI, but soon we realized that far more companies were eager to share their experiences and learn from one another.

Between 2023 and 2025, we conducted dozens of interviews with AI leaders at more than 20 companies and convened more than six events with representatives from more than 50 companies to understand how the applications of generative AI tools were evolving. The companies that participated in the research and working group events were primarily large, established organizations that are among the leading firms in their industries – they were not primarily AI companies or startups aiming to unseat incumbent companies through their applications of AI. Many of the companies we studied were the incumbents themselves.

While we recognized that we were only studying the early stages of a long story, we sought to gather the perspective of businesses over time. How did AI and HR leaders within large companies discover what generative AI tools could do, and how did they learn from their experience? What did the rise of generative AI look like from the worker's perspective – the people who began using these tools, and whose jobs were being affected?

At the outset we thought that the adoption and impact of these technologies would vary by sector, so we identified four categories of organizations that we thought would provide a cross-section of the economy: i) healthcare and life sciences, ii) retail, iii) finance, insurance, and real estate, and iv) manufacturing. We were interested in the different approaches to adoption in each of these domains, as well as the common threads that applied across industries.

We gathered data on the experiences of companies and their workers through several channels. We conducted interviews with senior leaders responsible for deciding how AI tools would be deployed, as well as the leaders charged with understanding how jobs and hiring would change due to generative AI deployments. Where possible, we spoke with individuals with direct experience using the applications of AI that had been deployed. Often the individuals we interviewed were managers within their organizations – not the frontline workers or individual contributors whose jobs have been affected by the implementation of AI tools.

We also draw on the results of a large-scale survey on workers' attitudes toward and experience with AI and automation conducted in late 2023, as well as additional survey data of the public on AI usage patterns. These survey data serve to complement the interviews with management. Where the survey data aligns with what we learned from management, we can treat the interview data with additional credibility. Where the survey data and management interviews differ – for example, if managers suggest workers are enthusiastic about an application of AI, but survey data indicates worker opposition – then we can highlight the variation in perspectives.

In addition to interviews with participating companies, the research team convened a working group of more than 50 companies and other organizations to meet periodically to discuss progress and challenges that AI leaders faced as they deployed generative AI tools. The meetings included five virtual sessions, as well as two in-person sessions, each of which had speakers from companies and other organizations who shared their work. Whereas interviews with organizational leaders and managers yielded detailed information about applications of AI, the working group sessions identified where organizational leaders saw consensus around the utility of these new tools, and where there were still unanswered questions. Throughout all these interactions with companies, we assured research subjects that they and their organization would remain anonymous unless they provide permission to publish their name. The organizations cited were given the opportunity to provide comments on the findings, but did not have any editorial control.

Across the research, there were several key themes that we sought to address.

First, we were interested in how the division of labor between workers and technology was changing. What were organizations asking generative AI to do, and how was the role of human workers changing as a result? This set of descriptive questions guided much of our qualitative data collection. The answer to this question would help us evaluate whether jobs seemed to be improving – and how dramatic the impact of these tools seemed to be on the labor market in the short term.

Second, we were interested in the shift in skills and training required to work alongside generative AI. Although it was often unclear which employees would be the winners and losers of this technological shift as employers were experimenting with generative AI, employers still had some sense of which skills would be important for their organizations into the future – and how they prepared to train for those skills. This question allowed us to examine the skill leveling hypothesis.

Third, we were interested in how companies decided which generative AI applications to scale up, and which applications to abandon. Since we were observing companies at their earliest stages of generative AI use, we observed significant learning over time within organizations as they identified potential applications of the technology before identifying which uses were most promising. How organizations managed the process of iterating on early applications of the technology allowed us to test the third hypothesis.

This section presents the findings on these three questions in three roughly chronological stages. In each stage, the findings highlight differences between the sectors' experiences with generative AI tools, along with common threads that we identified across the companies' experiences.

STAGE 1: TRY EVERYTHING

In late 2022 and early 2023, as employees within companies started using generative AI tools to help them with their work – drafting memos, summarizing documents, searching for information – there were two opposing reactions. The first was conservative. There were concerns that employees might be sharing proprietary information with the generative AI tools that could be exposed to other users. At multiple companies, there was an email that circulated to employees guiding them on authorized use of generative AI tools. The sentiment was as if the organizations were encountering an alien species: we don't quite understand what this technology is yet, so we need to be cautious about how we use it.

A second reaction was filled with excitement and ambition. Leaders within organizations had heard the hype about AI for years. Now it seemed that useful tools were suddenly available to the masses. As one AI leader within a large insurance firm put it, there was a feeling that this was an "iPhone moment." This was not the first AI technology that the organizations had tried, but the scale and usability seemed fundamentally different.

The possibility of generative AI technology represented both an opportunity for the organizations that we studied to grow their productivity, as well as a threat that upstart organizations – armed with this technology – may have the potential to "disrupt" their business advantage. For many of the organizations we studied, which were leaders in their industries and had been operational for decades (in some cases more than a century), the fear of being disrupted in the ways that Clayton Christensen outlined in *The Innovator's Dilemma* motivated quick action to experiment with generative AI tools.

TOP-DOWN VERSUS BOTTOM-UP IMPLEMENTATION

Based on previous research on how software tools are deployed, we expected a mix of top-down introduction of generative AI tools (following the priorities and instructions of CFO and CTO like figures), as well as a call for bottom-up input of how the tools should be used. The potential difference with generative AI tools was that they are available as consumer tools as well as enterprise tools. Therefore, companies had less control over the bottom-up uses of the tools. In some cases, companies in our sample felt pressure to adopt the tools at an enterprise level because they assumed that their employees would be using the tools anyway. And they wanted their employees to use the tools safely. This dynamic accelerated the pace of adoption and appeared to shift the pattern of adoption toward bottom-up use cases.

There are forces within organizations – often from the top – advocating for planning and patience. Typically led by a task force of senior representatives across an organization, they start by setting policy and recognizing how the new technology can align with the company's overarching strategy. The vision for how new technologies might transform the company is clear, but they are often deliberate in rolling out new tools across the organization. For example, multiple large companies we studied established a committee of senior leaders dedicated to ensuring that all employee interactions with their LLM align with their AI governance principles and contribute value to the organization.

And then there are forces pushing for speed and agility, the type of startup mentality associated with "failing fast." They recognize that the technology is evolving quickly, and there is no dominant design. But their goal is not to build an application that lasts for years. It's to figure out how these tools might be useful now – and adjust as they change and improve.

Many companies have built sandboxes for employees to use an LLM for their own purposes. With a little training and some guidelines on what types of use cases are permitted, companies can trust their frontline workers to identify the best ways to use Generative AI tools in their own jobs. The challenge is identifying the best ideas to scale. That's where learning by monitoring comes in – this approach requires a central team or mechanism to understand what's paying off so that participants can learn from one another.

We recognized in the working group companies three models for experimentation, which roughly correspond to a top-down, bottom-up, and hybrid (center of excellence) strategy. Bottom-up approaches benefit from the flexible nature of some generative AI tools. The companies we studied frequently used some version of the bottom-up approach, but the value of this adoption strategy is hard to quantify – and it is not clear how this adoption can lead to deeper, more valuable uses that enhance productivity. Top-down approaches can identify deeper use cases that will add value for the organization, but they require overcoming potential grassroots obstacles to adopting the tools (in contrast to the bottom-up dynamic).

It is important to note that many organizations we have observed are running some combination of these approaches.

For example, Lambda is a Fortune 500 insurance company. Years before adopting generative AI, the company had grown a large data science team that had developed AI applications to improve their customer relationships and claims processing for years. In early 2023, as the company the data science team was charged with creating a "sandbox" for the insurance company's employees to use generative AI using the insurer's own data in a secure environment.

The speed with which the company mobilized to create the internal platform was remarkable. The directive came down in April 2023, and the executives supporting the project wanted the platform ready by May 1. "I've never seen a technology go from nothing to being in production as fast as this has," a Lambda data science leader told us.

Soon thousands of their employees were using the platform. Little training was required – just a 10-minute tutorial on how to use the tool securely. The idea was that their tens of thousands of employees would imagine a wide variety of uses for the internal generative AI tool. Lambda data scientists did not have the expertise to anticipate how their colleagues could best use the generative AI tools, so they harnessed the flexibility of generative AI tools and provided a platform for individual users to experiment with the applications of generative AI that would work best for them.

There have been many companies like Lambda that have supported grassroots experimentation with generative AI tools among their workforce, often with internal tools that modify or otherwise resemble ChatGPT, Claude, or Gemini. The promise of letting a thousand flowers bloom is to identify use cases of the technology that senior executives and software engineers could not have imagined on their own.

At the same time that Lambda was facilitating bottom-up applications of generative AI, senior leaders within the organization were working with managers to identify use cases of generative AI that they could apply to key tasks within the organization. The first step in this process was to solicit ideas on use cases based on the priorities for the company. Then Lambda would present the use cases to a "responsible AI committee" to review the use cases, their potential financial impact, and how they aligned with regulations and the organization's AI principles.

The data science team noted that more than 300 use cases had been proposed to the AI team in 2023. A far smaller share – around thirty – were selected to become live generative AI applications. This process – first soliciting ideas of all the tasks that AI could accomplish, then culling the list to identify the highest-impact uses – was common to many of the organizations that we studied.

This top-down approach is consistent with how software tools have been applied in the past. For software tools like robotic process automation, organizations have sometimes developed scorecards that describe the use case and estimate its return on investment before a central committee can review whether it is worth pursuing.

Figure 1
Two Approaches to Experimenting with Generative AI

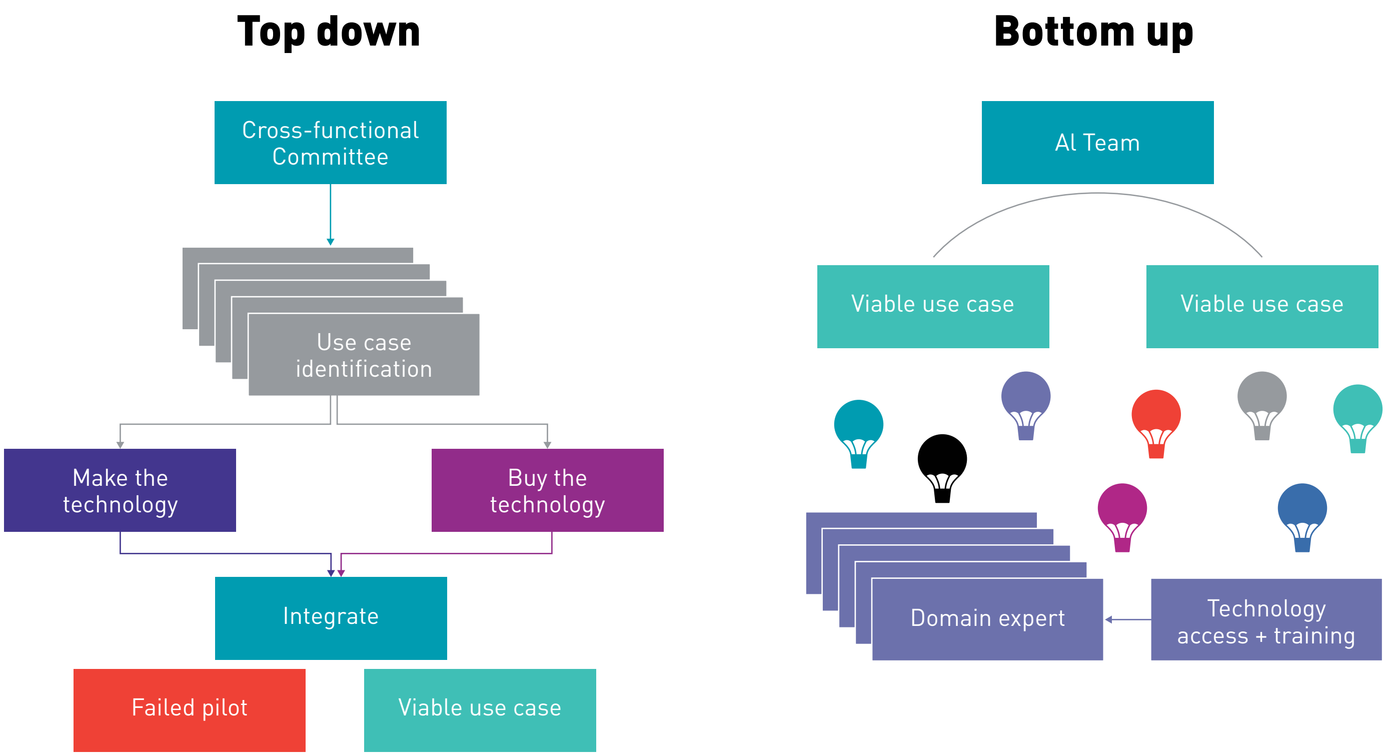

Figure 1 outlines the rough steps in each of the top-down and bottom-up approaches to identifying the problems that generative AI could help solve.

THREE PROBLEMS

Across the more than 20 organizations we studied experimenting with dozens of unique applications, the uses of generative AI were directed toward several common problems. Defining these common problems provides a structure for understanding the motivations behind a diverse set of industry-specific applications of generative AI. They will also be helpful in determining how organizations can define the role of the human in the loop as they reshuffle tasks between human workers and AI tools – and redesign jobs as a result.

The first is the bottleneck problem, which refers to the growing volume of near-routine tasks that require skill to perform, but prevent domain experts from working on more complex and interesting work. Addressing the bottleneck problem aims at relieving professionals of tedious tasks that are frequent enough to be time-consuming and not routine enough to automate with previous technologies. Spending time on these tasks – which workers can find burdensome – can hamper workers' productivity because they are spending time on work that adds less value to their organization than if they were fully using their expertise. When generative AI tools address the bottleneck problem, they often shift the worker from executor of the near-routine task to supervisor of the task, reviewing the output of the generative AI tool.

The second challenge is the cafeteria problem, which is the challenge of integrating information from diverse experts across an organization into a coherent output, whether it be a product, document, or idea. We refer to this as the cafeteria problem since organizations said that the current way of addressing this challenge would be to go to the cafeteria and find the person with relevant expertise to ask them to weigh in on a particular question. The promise of generative AI applications in addressing this challenge is to synthesize the documents and other information from within an organization – alongside information from public sources – to provide a prediction of how experts within an organization would be most likely to respond to a query. The purpose of the generative AI application would be to empower the team focused on integrating information from experts – and save the domain experts time in responding to the queries.

The third challenge that organizations have sought to address with AI is the learning curve problem. Tasks that require a high level of expertise – or knowledge of a complex domain – can take a long time and require personnel who are in short supply. One challenge is that training new experts in complex fields can take a long time and require incumbent experts to take time away from their roles to provide training. The opportunity that organizations envision for generative AI is to provide responses to novice inquiries that draw on the extensive documentation and data that experts have already internalized. Generative AI tools that help novices climb the learning curve may also help experts extend their knowledge by accessing new knowledge in their field. For tasks that have more available information than one individual can possibly retain, the generative AI tool is designed to help individuals access either A) more knowledge than they could have otherwise, or B) the same knowledge that they would have otherwise, but in less time.

Each of the problems that organizations have looked to generative AI to address are defined at the task level – not the job level. It may seem obvious, but organizations are not aiming for AI tools to assume all the tasks of any job. Instead, they are targeting these applications to reduce the time that workers take on certain tasks or increase their performance on others to improve the productivity and quality of human teams overall. Despite common discussion among some AI companies and the popular press about AI displacing certain jobs, the vision and implementation of early AI applications has frequently been to automate or augment individual tasks or small groups of tasks, such as writing reports, analyzing documents, or programming software.

Understanding how organizations across industries have addressed these common problems works to highlight the common threads among early applications of generative AI, as well as the evolving role of workers across industries.

HEALTHCARE AND LIFE SCIENCES

At the leading hospital systems that we studied, there were two types of problems that generative AI applications set out to solve. The first was focused on reducing the administrative burden for doctors and nurses. After the COVID-19 pandemic, hospitals faced higher turnover from doctors and nurses due to burnout.

One factor that departing medical professionals cited that was hurting their job quality was the amount of paperwork required to do their jobs. They spent a growing amount of “pajama time” maintaining patient records and responding to electronic inquiries about patients’ health. While maintaining patient notes to ensure that medical visits can be reimbursed from insurers is a longstanding practice, the response to patient messages online is relatively new.

During COVID, more patients became comfortable asking their doctor a range of questions about their health, leading to a surge in inquiries. Now doctors and nurses had to manage an in-person practice *and* respond to a hotline of real-time questions. The challenge for relieving the administrative burden was whether new technologies could save medical professionals time in communicating with patients and maintaining their records without sacrificing quality.

The Mass General Brigham Healthcare System, a leading research hospital network, experimented with two applications to relieve doctors’ administrative burden. The first was a generative AI tool that supported response to inbasket messages. The tool, which integrated into the hospital’s electronic medical record system, provided doctors and nurses with a draft response to a patient’s message. The goal was that rather than draft the initial response themselves, doctors could merely edit the LLM’s response and manage the body of patient messages more quickly.

A second application sought to assist with the preparation of medical notes for patient visits. Previously the hospital system had experimented with recording patient visits and sending the recordings to a third-party office of medical professionals (outside the United States) to transcribe the medical note and annotate it with a relevant code so that the visit could be reimbursed. The doctor seeing the patient could then review the transcription of the medical note for accuracy and approve it.

The application of generative AI envisioned a LLM-based tool that could take over the role of the third-party, automatically transcribing the medical visit recording and summarizing it into a note with annotations for the doctors to review.

Aside from these two applications aimed at reducing healthcare providers' "pajama time," there were other use cases that focused on improving the quality of medical care. A second hospital system developed a LLM-based application to summarize patient notes more quickly in the case of handoffs between shifts for nurses. They had found that a high-quality handoff was important for avoiding medical errors – and that the handoffs were taking so long that nurses would frequently arrive home late, and the extra time was reducing their job quality. The time-saving generative AI application was aimed at decreasing the time burden on the one hand, and ensuring quality on the other.

It is worth recognizing that these applications of generative AI tools have evolved in parallel with extensive use of narrow or traditional AI and machine learning technologies in healthcare, most notably for assisting with diagnostics. The most prominent example is in radiology where medical systems have sought to integrate AI tools for supporting improved diagnostic accuracy for the detection of cancer and other illnesses. Individual doctors using generative AI for assistance in conducting research related to their work. However, the use of these tools for diagnoses appears to vary from individual to individual without a concerted effort at the organizational level to augment the diagnostic or research tasks of the doctor with generative AI.

In the life sciences, we were interested in studying the top problems that pharmaceutical companies and related organizations set out to solve with generative AI. Like large hospital systems, these companies had experience using AI for research and development purposes. Some of the most prominent and celebrated early applications of AI tools have been for drug discovery.

In our research, pharmaceutical companies like Bristol Myers Squibb (BMS), a large inventor and manufacturer of medicines and therapies, focused their early applications of generative AI on reducing the time to win approval for new drugs. Gaining approval for a new treatment from regulatory agencies is often the bottleneck in the process for companies like BMS. If they could make the process faster by weeks or even days, patients would be able to access treatment faster – and there would be significant financial benefits for the company. The cost of shepherding a drug from discovery to approval would drop.

The opportunity they identified first was not in accelerating the research process, but the paperwork required to show that the proposed drug could pass regulatory hurdles. Historically, once a clinical trial was complete, an application to the NDA could take as much as 12 months. As an industry, they have shortened the timeline to near 6 months. BMS's ambition was that with generative AI tools they could shorten the time to 3 to 4 months.

The tool to achieve that kind of speed would pull information into the application without requiring the BMS team to reach out to diverse experts – or access information in hard-to-find locations – and assemble it all manually. This was the equivalent of tracking down experts in the cafeteria to gather their input. The ambition was that generative AI tools could anticipate what the experts would contribute based on the internal documentation that had already been prepared in the course of years of research, clinical trials, and other internal documentation.

BMS recognized that if they could figure out how to augment the documentation process in clinical trials, they could apply the same kind of tool to other types of documentation, too, reducing the manual burden of assembling the perspectives of diverse experts. The challenge, of course, would be whether they could guarantee the quality of the documents that they produced if an algorithm was playing a role in putting them together. That's where they needed to consider the role of the human in the loop, redesigning the job of the person charged with managing the documentation process. The user of the proposed generative AI technology would still need to be a domain expert who understands what good responses to the compliance process would look like.

Table 1
Early Applications of Generative AI by Industry

No.	Industry	Challenge	Vision for generative AI application
1	Healthcare	Medical professionals face high administrative burden due to soaring patient messages **(Bottleneck problem)**.	Produce drafts of messages in response to patient inquiries for medical personnel to review and modify.
2	Healthcare	Medical professionals face high administrative burden due to maintaining patient records after visits **(Bottleneck problem)**.	Transcribe patient recordings and summarizes the transcript in the proper format of a medical note for medical personnel to review and modify.
3	Life Sciences	Applying for FDA approval and managing drug development process is paperwork intensive and requires complex coordination of domain expert **(Cafeteria problem)**.	Produce draft language for sections of the application paperwork based on internal documentation and LLM training for internal personnel to review and modify.
4	Finance	Responding to client RFPs is paperwork-intensive and requires complex coordination of domain experts **(Cafeteria problem)**.	Produce draft language for sections of proposal based on past reports, internal documents, and LLM training for internal personnel to review and modify.
5	Insurance	Field personnel evaluating assets to insure requires comparing individual contexts to complex standards and regulations **(Learning curve problem)**.	Responds to field personnel questions and notes by summarizing relevant internal documents that are relevant to the questions posed and individual circumstances.
6	Real Estate	Agents' comparative advantage in serving clients is up to date market information **(Learning curve problem)**.	Responds to real estate agent inquiries with customized data based on internal documentation.
8	Retail	In-store personnel face a wide variety of questions and often have limited information to answer them **(Learning curve problem)**.	Responds to in-store personnel with customized response based on internal data and documentation.
9	Retail	The success of products sold online can vary based on titles and product descriptions **(Bottleneck problem)**.	Produces draft product titles and descriptions for review; narrow AI can analyze which types of reviews are most effective.
10	Manufacturing	Building new products often requires integrating knowledge of diverse engineering fields to get started **(Cafeteria / learning curve problem)**.	Coding support for mechanical engineers without software expertise to program mechatronic products using Arduino.
11	Manufacturing	Factories frequently have diverse equipment, each with complex manuals, that require specific expertise to maintain **(Learning curve problem)**.	Responds to technician inquiries about how to troubleshoot problems that arise on equipment in a production environment.
12	Multiple	Software development faces growing demand; creating and testing new software tools can be time-intensive and often requires knowledge of complex syntax rules **(Bottleneck problem)**.	Drafts code for software developers (and novices) based on natural language prompts; can also run unit tests and provide instructions for deploying code.
13	Multiple	HR representatives are often presented with routine, easy-to-answer questions that have small variations based on an individual's circumstances **(Bottleneck problem)**.	Provides automatic responses to personnel asking routine HR questions; routes more complex questions to HR personnel.
14	Multiple	Customer service representatives are often presented with routine, easy-to-answer questions that have small variations based on an individual's circumstances **(Bottleneck problem / Cafeteria problem)**.	Provides draft replies to HR representatives based on inquiry; OR automatically answers customer inquiries that are routine, routing more complex questions to representatives.
15	Multiple	Lawyers and clerks have time-intensive document review tasks that require identifying slight variations in language across a large corpus of documents **(Bottleneck problem)**.	Reviews documents and identifies or extracts key information based on lawyer inquiries; summarizes documents based on information of interest.

FINANCE, INSURANCE, AND REAL ESTATE

MFS is a large investment firm with billions under management and more than a century long track record. The firm's value to its clients, which are primarily large institutions that choose to invest with MFS, is its knowledge of market trends, as well as its trustworthiness. MFS has long been an established firm within the industry. It began experimenting with generative AI applications by appointing a small committee representing different groups within the firm to identify promising use cases for the technology. The financial industry was not new to AI tools – there was a precedent of using machine learning techniques to analyze the market and determine optimal investment strategy.

Generative AI, however, represented a different kind of opportunity. The team was aware that generative AI could synthesize knowledge and compose essays and code, but they were also cautious about the potential for generative AI tools to produce errors that might affect their clients' trust.

One of the first applications MFS explored was for generative AI to assist in the development of proposals for new clients. The firm won new business by responding to requests for proposals (RFPs) and due diligence questionnaires, which required an internal team to reach out to experts throughout the business to understand the firm's capabilities and knowledge across different domains that might be relevant to the proposal. It was a labor-intensive process that often required a small share of time from many people. It's another manifestation of the cafeteria problem that BMS sought to solve.

The vision was that a generative AI tool could speed up the process by generating an initial draft of the firm's response to the RFP. The idea was to anticipate how the firm's experts would respond, matching content the firm had already approved with the questions from the RFP. Then, the firm's internal team could edit and improve the document. A successful application of generative AI in this case would continue to require a domain expert to review the output of the AI tool with the knowledge and skills to assess its accuracy and appeal to potential business partners.

Epsilon Insurance provides insurance for valuable assets. Like MFS, its business relies on expertise and trust. Also similar to MFS, Epsilon's approach to experimenting with generative AI started with a committee of senior experts within the organization that filled an extensive spreadsheet with potential use cases within the firm. Epsilon's expertise comes in valuing the risk associated with different assets – and even advising customers on how to reduce potential risk. Their field agents often have technical backgrounds and must undergo extensive training to ensure that their assessments follow extensive manuals and guidelines. They must also prepare documentation related to their assessments of the property that they insure.

The vision for generative AI at Epsilon is to provide assistance to personnel in the field: responding to their inquiries to access additional knowledge, as well as support in documenting the work that they do in the field. The goal is to shift the tasks of the field agent from actively searching and reviewing documentation in order to make judgments about the risk of a property – to reviewing information and assessments provided based on a combination of data internal to Epsilon, as well as from context provided by the field personnel and public information.

This application of generative AI represents a combination of the learning curve problem, as well as the bottleneck problem. Access to knowledge from the generative AI tool could help build expertise for junior personnel, in lieu of leaning on senior experts. It could also assist in the time-intensive tasks of risk analysis for field personnel that do not depend on their expertise, but still require attention for purposes of formatting and compliance.

The role of the field personnel would shift from actively searching for, reviewing, and synthesizing information – to reviewing the information that has been synthesized on their behalf for accuracy and relevance. The challenge of these applications, as in the case of MFS, is ensuring the accuracy of the information that the generative AI tool provides.

Gamma is a commercial real estate company with two main business streams: brokerage and property management. Gamma's approach to experimenting with generative AI was structured. They approached the introduction of the technology as an opportunity to upgrade their technology capabilities in what was often considered a legacy industry that relied on generations-old technology tools. The technical team at Gamma envisioned that generative AI applications could take lagging software systems and unstructured data infrastructure – and improve it substantially.

Gamma's initial focus was providing resources to its real estate brokers. In the industry, real estate brokers are high-value assets, since they are free to shift between firms and their relationships and expertise are associated with substantial added value for the firm. One of the main resources that a firm like Gamma can provide to real estate brokers is data and technology that can make it easier to find the right property fit for clients – and can reduce the transaction costs when properties are bought, sold, or leased. Gamma realized that it would not be building all the technology tools with generative AI for its brokers, but it wanted to provide the right resources to help them become more effective.

One problem that Gamma sought to solve for its brokers was efficient access to market data that was customized to their inquiries about particular geographies and clients. Gamma maintains proprietary data and draws on public sources to provide a real-time view of the real estate market. However, it had traditionally been difficult to access these data without the technical expertise necessary to query a database and pull a specific statistic. The workflow in Gamma, like many business service organizations with a sales team, was often that a broker would need to request the data from Gamma's technical team, which would have to assemble it. The vision was that generative AI could allow brokers to query the data on their own in natural language, avoiding a complex – and potentially time-consuming – back and forth with a data science team.

The Gamma application for its brokers is an example of the learning curve problem, where the application is designed to extend the capabilities of its brokers – already well-versed in client relationships and experienced in market dynamics – with access to data against which they can test their assumptions and expand their knowledge. It is unclear, however, what skills are required to make successful use of a generative AI tool like this one. For instance, brokers may still need to understand the underlying patterns in the market to interpret the data that a generative AI tool provides.

RETAIL

There were two retail domains in which the organizations we studied sought to deploy generative AI: online and in brick and mortar stores. Tau is a large retailer with a large e-commerce business, as well as hundreds of brick and mortar stores. The company has invested over the past decade in substantially expanding its software and data science teams, which prepared it to develop internal capabilities once generative AI tools became available.

Tau's applications of generative AI for e-commerce were aimed at a bottleneck problem. Tau frequently added new products for sale on its e-commerce website. Each of those products needed a title and description. Those title and descriptions were particularly important because the retailer needed to ensure that if a customer searched for a product on the internet, the retailer's listing of the product would show up near the top of the results.

Tau started by passing each of its product descriptions and titles through a customized generative AI application developed to ensure that each of the products would be most likely to appear high on search results. As their technical team became accustomed to customizing generative AI tools for use on their website, they saw additional opportunities, including one that would allow customers to ask questions about the product with a generative AI enabled chat tool on the retailer's website.

This was a new capability for the retailer. The product pages each had frequently asked questions, as well as their answers from user comments. However, the advance that generative AI offered was for customers to ask specific questions tailored to their own needs. If a customer was searching for a kids' bed for their ten year old, the customer could ask whether the bed would fit a ten year-old. The generative AI chatbot would respond based on the product description and manual, if available. The assumption was that the chatbot could respond based on the expected size and weight of a ten year-old, as well as the size and weight that the documentation for the bed considered appropriate.

Tau's primary application of generative AI within its stores sought to refine an existing technology tool, rather than build a brand-new capability. For years, the company had sought to organize the retailer's data for in-store associates in an accessible fashion via an app on tablets that were available to in-store associates. However, the information that those associates often received from the company's app (before generative AI) were not useful. This had been a common problem for company-specific search engines that point users to links to resources rather than provide answers to questions.

Generative AI seemed to have capabilities that unlocked the potential of the app for in-store associates. A chatbot enabled with generative AI that had access to the retailer's data could provide useful answers to an associate's queries. The goal was to address the cafeteria problem. To answer customer questions that they could not answer on their own, an associate would typically need to contact a manager or domain expert for assistance. If they could get an answer on the app, then they could provide better customer service (by responding more quickly) and preserve the time of their colleagues.

What they also found, after deploying the tool, was that employees would pose questions that the technical team behind the tool had not anticipated. Rather than merely asking about store inventory or information about a product, employees would also ask personal questions that revealed their insecurities on the job – or dynamics with their team. It appeared that this tool designed for productivity and decision support was also a channel for workers to provide feedback about their work experience to their managers that they may have never communicated otherwise.

MANUFACTURING

Although early generative AI applications were associated with high-skilled office work, manufacturers face many of the same challenges that financial firms and healthcare organizations do, albeit in a different environment. Building a prototype of a new product requires significant documentation and diverse technical expertise that must be integrated. There are also plenty of near-routine tasks on a factory floor that a machine could perform – as well as additional technical knowledge that could be beneficial to novice technicians.

The manufacturers we interviewed saw the potential for generative AI applications in several domains, but were skeptical of the potential of these tools to meet the quality standards in a manufacturing environment. Factories had adopted some AI tools, including computer vision to conduct quality inspections, but most of the new technologies in manufacturers came from outside vendors. The software expertise to modify new generative AI programs within manufacturers was limited. Nonetheless, there was some experimentation from the bottom-up.

For example, at the large diversified manufacturer Phi, a mechanical engineer worked with a team to develop mechatronic prototypes that they could modify and improve before producing them at scale. The engineer's key contribution was in the mechanics of the product, but the electronics needed to be passable. The engineer had previously tried to teach themselves how to program in Arduino using online forums. The process was halting and frustrating. As they tried to learn, those who responded in the forums were rude and belittling. A generative AI coding tool was a welcome alternative, answering his questions with a cheery disposition. The engineer's use case became a model within the firm.

Other manufacturers envisioned a generative AI application that would solve the learning curve problem, similar to the insurance and real estate industries. Manufacturers in the United States, in particular, have a bipolar workforce: a growing share of young people, and a concentration of technical experts soon to retire. A generative AI tool that can help novice technicians troubleshoot problems on the factory floor or anomalies in the production process – which draws on the manuals and documentation for machines – could boost the effectiveness of inexperienced technicians faster than they would have if they learned through on-the-job observation and mentoring.

Technology providers for the manufacturing industry have suggested numerous applications of generative AI in manufacturing, including robots that are prompted to perform tasks with natural language, as well as design tools that generate usable designs – and usable machine code – based on existing drawings. While these ideas reflect potential directions that technology providers find promising, these applications are not clearly linked to existing problems that manufacturers face – or the common challenges that generative AI has been envisioned to solve in other industries.

COMMON APPLICATIONS

Across the industries represented in the working group, we noticed that common use cases were just as frequent as industry-specific ones. Hospitals, financial institutions, and technology firms were all experimenting with uses of generative AI in HR, customer service, software, and law. These common use cases tend to respond to similar problems as the industry-specific use cases. The primary difference, at least at this early stage of generative AI experimentation, is that there are numerous technology companies developing AI tools for legal, customer service and software use cases, for example, which have only served to accelerate the use of AI in these common areas.

Software development appears to be the leading common application of generative AI across firms. When generative AI firm Anthropic released data on the estimating the usage of its generative AI tool Claude, more than one-third of the usage appeared to be for "computer and mathematical" tasks, which are likely related to software development (although workers in this category comprise just 3% of the workforce).

Among the working group organizations, we heard repeated examples of software development teams integrating generative AI tools to improve the speed with which they could prototype new ideas. At the retail organization that developed the generative AI application for their in-store associates, they were given just six weeks to turn around a prototype – a much faster timeline than they were accustomed to following. As our interviews progressed, companies began discussing AI adoption for computer programming as a *fait accompli*. It was assumed that their software engineers were using these tools. However, since many of these companies do not sell software, the question was often how they could use these tools to generate value for their core business.

Organizations have almost universally presented the generative AI software tools as helpful, but their accounts of the benefits vary. Some software engineers report being able to produce a week's work in a fraction of the time – a classical case for productivity improvement that might suggest software teams will require fewer people. However, other organizations have suggested that the time-savings will allow them to take on projects that had been in their backlog. There is no shortage of new technology projects to address, they argue, and these tools allow them to take on a greater breadth of work.

There is also variation in how organizations think about the problem that AI-enabled software engineering tools are addressing. For some software engineers, like Mark from the introduction, software engineering tools have helped address the bottleneck problem: they are taking over near-routine tasks, like unit tests, that had been frustrating. For Mark, this represents a logical evolution in software engineering, where the practice has evolved from writing machine code to increasing levels of abstraction made possible by APIs and extensive open source code libraries. For some, AI code assistants are the next step in this evolutionary trend.

Early reporting on the de-skilling of software engineering and the rise of "prompt engineers" suggested that code assist tools were addressing the "learning curve" problem, making it easier to work like an experienced computer programmer with less training. Although we saw limited bottom-up evidence of this activity in manufacturing – the coding for mechatronics example – this was not the prevailing use case for the companies that we studied.

Organizations across industries also developed applications for human resources and customer service teams. These applications were often aimed at the bottleneck challenge, where HR associates would receive similar versions of a common question (from employees) and customer service representatives would receive similar versions of a common question (from customers), but it would require a human to categorize the question and offer the common response. The vision in both cases was for generative AI to categorize and respond to these common questions, leaving human representatives to handle the harder-to-answer queries.

It is worth noting that a generative AI application to solve the bottleneck problem would need to be designed differently than one aimed at the learning curve problem. The goal of the application focused on the bottleneck would be to respond quickly to near-routine queries. The human in the loop would be expected to know how to categorize those queries. The learning curve application would require more engagement from the human in the loop as the generative AI tool suggested potential responses and resources to address more complex questions.

In law, the primary application we observed was in response to the bottleneck problem. Generative AI has been used to assist with document review – finding similar, but non-identical clauses in a large corpus of documents that would be time-intensive to review manually – as well as for drafting documents that draw on common clauses or a consistent structure. In these cases, as the others, the vision is not to automate these tasks, but to shift the lawyer from a role of manually performing the tasks to reviewing information that the generative AI application has processed.

In marketing, the bottleneck was that customizing communications for individual audiences across communications channels took substantial time. A life sciences company Theta developed an internal generative AI application to address this problem. It looked like a web 2.0 game. You could give it a small blurb about a product, an event, or other piece of information. With a few clicks as if filling out a survey, the application would create a marketing campaign – tweets, emails, posts in a variety of tones and for various audiences. The tool became one in a family of generative AI applications that the marketing team had built – not just to create copy, but to perform research, support podcasts, and more.

STAGE 2: MEASURING SUCCESSS

When we first spoke to many of the organizations in the study, they were just beginning to imagine how generative AI tools could improve their work. In some cases, the use cases were still just a vision and had not yet been prototyped. Among the organizations that had live use cases, they frequently measured the success of those early products in terms of how many employees were using the tool. Attracting early users, it seemed, proved that there was demand for the application and the use case was viable.

Then came the revenge of the CFO. After roughly the first year of experimentation with generative AI tools where technology teams had extensive budgets to try new applications, there was a wave of belt tightening. Usage was necessary, but not sufficient, to justify investment in new generative AI tools that often required expensive contracts with technology providers and expensive technical expertise.

This shift paralleled a transition from prototyping to production: if the organization would invest in and scale up the use of generative AI tools, then they would need to measure the return on investment. Four themes from this period of generative AI experimentation cut across the organizations that we studied.

I. BOTTOM-UP AI APPLICATIONS PROVIDE FREEDOM TO WORKERS, BUT THE BENEFITS VARY WIDELY, MAKING THEM HARD TO MEASUREE

It is hard to measure the impact of individualized use cases across an organization when everyone can customize how they use the tool. Some uses of the tool may be highly valuable, while others may be distracting or lead to costly errors. The dilemma of bottom-up applications is that when organizations invest in giving their workers flexible technology, there is a leap of faith that the workers will figure out how to use the tools successfully.

This is a familiar problem. When companies adopted personal computers and the internet, there was recognition that the technologies could be transformative – and there was clear evidence that individuals could perform certain tasks more productively. Consider sending an email versus sending a fax. And there were new tasks that workers could perform that they previously could not, such as complex calculations. Nonetheless, the adoption of these tools and their impact on the aggregate productivity in legacy industries was slow.

Research on the "productivity paradox" documents how – despite widespread adoption of IT tools in businesses – productivity growth did not change substantially during the computing revolution. One theory is that productivity does not accurately reflect the benefits of digital technologies for improving product quality and driving down prices. But another explanation is that while IT technologies enabled workers to perform some value-added tasks faster, they may have also distracted workers to spend a higher share of their time on tasks with less value-added.

We did not find in the organizations we studied a method for assessing the value of bottom-up generative AI tools for their business. Instead, the attitude seemed to be that subscriptions to generative AI tools were an operating expense along the lines of providing employees access to the latest computing hardware or databases: these were tools that would enable their workers to perform at their best.

II. THERE ARE MULTIPLE PATHS TO HIGHER PERFORMANCE: SAVING TIME; IMPROVING QUALITY; AND CREATING NEW CAPABILITIES.

After organizations celebrated the usage of generative AI tools, they began highlighting the time savings that these tools provided. A frequent comment we would hear from organizations was that adding generative AI reduced the time to perform a given task by X%, or Y hours. Since productivity is measured in value added per labor hour, fewer hours meant higher productivity. The question often followed: what did employees do with the hours saved?

The focus on time savings often overlooks an alternative channel for generative AI tools to improve performance: improving the quality of the job that workers do, even if it takes them more time. For example, in a study of the impact of generative AI tools on doctors' ability to diagnose patients, the group of doctors using generative AI tools to assess patients took a longer time, but their diagnoses were judged to be higher quality.

Across the examples from organizations we studied, time savings were but one of multiple potential desiderata. Improving the quality of work – or gaining new capabilities – were as important, if not more so. They were also sometimes easier to measure.

For example, organizations like the retailer Tau recognized that generative AI could help address a longstanding learning curve challenge that previous technologies had not sufficiently addressed. It was not easy to measure whether having a generative AI chat assistant on a tablet boosted the value added that each store associate generated per hour – the classic measure of productivity – but it was possible that if the customer service representative was able to respond more quickly to a customer's question, they would be more likely to purchase an item at the store – and come back to the store with the expectation that they would find what they needed.

As organizations processed the lessons from their initial experiments, we heard consistently that productivity via time-savings was just one measure among multiple. One manufacturer emphasized that they had stopped looking at productivity gains from generative AI applications because the impact of individual technologies on productivity were too hard to measure. There were many factors that affected the firm's productivity, from the demand signal for their product, the skill of the team, and supply chain forces outside their control. Where they thought they could measure the benefits of generative AI was in the new capabilities it offered their workers.

For example, AI tools for computer vision in manufacturing allow manufacturers to gather far more image data on their products than they could have with previous generations of technology. One potential benefit of computer vision for quality inspection is productivity gains: whereas a human might need to visually inspect a part, the computer might be able to make the inspection faster. However, the apparent gains from this shift can be deceptive: the human in the loop doing the inspection often has other tasks in the production process and cannot be removed entirely. Moreover, the visual inspection might be very quick – even faster than the computer vision program. Some manufacturers in regulated settings might still want a human in the loop doing some visual inspection alongside the vision model.

Even if these limitations of the computer vision system mean that it is not a big productivity boost, the system itself provides the factory a new set of capabilities: the image data can allow the company to identify patterns in quality defects and potentially trace them to issues on the line – and companies with a large corpus of image data on their own products have more capability to use other AI tools to inform future process improvements and design adjustments.

There is a parallel application in the finance and insurance industries. For auditing, whereas manual audits would review only a sample of data from a vast corpus for an organization, the computing and analytic capabilities of AI allow auditors to identify anomalies across the entire corpus. Although this is unlikely to be time-saving overall, the new capability could lead to higher-quality audits that change firm behavior.

III. SHIFTING FROM BUILD TO DIY OR BUY

As generative AI experiments matured, organizations relied less on their internal IT teams to develop custom generative AI tools – and became more likely to buy generative AI solutions off the shelf, or to work through domain experts like consulting firms. In the first stage, we encountered organizations that tried to build their own custom AI applications (using fine-tuning or RAG techniques to customize one of the major LLMs).

Organizations held internal hackathons and other events to learn to develop new tools with generative AI. For example, a staffing organization built a tool that – with just a few parameters – would automatically generate a job description. The organization recognized that an application similar to this would likely be available to purchase from an AI company in due time – and they could buy it cheaper than it would cost to make the product on their own. Nonetheless, they opted to build it on their own to develop the experience and capability to work with generative AI tools.

Over time, the willingness to spend internal technology resources on efforts like these seemed to wane among the sample of organizations we studied. There were still technology teams internally that focused on deploying generative AI applications for priority use cases within the firm. However, as the number and variety of generative AI tools became available for purchase and licensing, the burden on those teams to build their own applications from one of the main large language models was lighter. As the expertise in AI development concentrated among emerging companies developing AI expertise, there was less pressure on the users of those tools in the companies that we studied to develop their own deep expertise.

There were exceptions to this trend, however. One technology company that had developed AI systems to manage its call centers and customer support pipeline continued to develop its own technology, rather than use technology from a vendor. The company provides office equipment and related software to enterprise clients globally. As they developed the AI tools for their own business, they began exploring if they could become a vendor of AI software solutions to their existing clients: the investment in making their own AI tools opened up a new capability and potential business line.

Comparing the modal organization – which opted to buy over build their own AI tools – to the organizations that pivot to develop and sell AI solutions highlights a fork in the road for organizations at this early stage of experimentation. Should they build up technology expertise to work with AI among their technical teams, or should they continue to rely on their domain expertise in their industry? The glut of targeted AI applications made available in the first two years of the experimentation with generative AI pushed many organizations to the latter road.

IV. AI IS NOT A FREE LUNCH

Yet another lesson from this stage of experimentation was the tradeoffs associated with the rapid adoption of generative AI. Of course, organizations recognized early on that applications using generative AI could produce inaccuracies. Embarrassing instances of generative AI use began to make headlines, including the story of a lawyer whose argument cited a non-existent case. The tradeoffs that organizations encountered in the second stage were both familiar and new.

New technologies that appear to deskill a task have long led to fears of skill atrophy. Just as there may have been concern about sewing machines prompting the decline of craftsmanship or internet search engines "making us stupid," there has been similar concern about integrating generative AI tools in ways that lead to useful skills disappearing from the workforce.

Early research on the impact of generative AI tools on individuals' ability to recall their tasks underlines the potential problem. In a study of undergraduate students performing a research task, those using ChatGPT had far less brain activity and ability to remember their work compared to their peers using a search engine or their memory. The implication was that the use of generative AI might yield passable results, but may not lead to skill development.

This possibility has significant implications for generative AI applications targeting the learning curve problem. Are these applications helping inexperienced workers perform as if they were experienced without gaining relevant expertise, or are they actually helping the novice users learn and gain expertise. This is a critical question for the value of generative AI tools in these domains.

A second, less familiar problem is the impact of generative AI tools on teamwork and mentorship. One financial organization reported to us that they were concerned the rise of generative AI use was associated with less collaboration among teams. One possible mechanism is generative AI used to address the cafeteria problem. Although the process of gathering expertise from diverse people within an organization might take valuable time from everyone, it is also a social process that could lead to shared understanding and learning that helps the team perform other important work in the future. Transforming the task of completing a project from a cooperative one where multiple people must pitch in to an individual one where the generative AI user is burdened with the work could have hidden costs. Not only may the team become less cooperative, but the individual job of completing the project could be less desirable as a solo effort.

A third tradeoff involves how new generative AI tools are developed. In the first stage, organizations faced the question of how to identify problems for generative AI to address. There was a bottom-up approach, as well as a top-down one, each with benefits and limitations. Once the problem was identified, there was the question of where to recruit the early users from within the organization to participate in the experiment.

The organizations we studied primarily focused on identifying generative AI "champions" or power users who were most excited to use these tools. Based on early trials with these users, the organization could gather data on the utility of their applications and refine them with feedback on the technology's usability. But several organizations recognized the limitations of this approach. Although power users are likely to give extensive feedback, their use of technology – and their workflow in general – may not be generalizable to the majority of workers in their organization.

For Gamma, the real estate company, piloting their AI applications designed for brokers by starting with the highest-performing brokers didn't work. Part of the reason was that the brokers with high levels of experience and extensive contacts needed different capabilities from a technology than brokers who were less experienced and were still developing their knowledge base. These less-experienced brokers also appeared to have a lot to gain from support from an AI tool.

Recognizing this challenge, organizations like Gamma began identifying the need to develop AI applications that were tailored toward "personas" of workers within their organization. One AI application would not fit all. Users skeptical of generative AI might respond better to applications that showed quick benefits without demanding much time, whereas power users were willing to spend more time engaging with the AI tool.

STAGE 3: SCALE AND SHELVES

After a wave of experimentation, organizations settled into roughly three modes of AI usage, each with its own potential impact on work. There were the organization-level applications of AI that were driven by problems that teams within the organization faced. There were also individualized applications of AI, which included the AI platforms that organizations made available to their workforce that they could use as they wished. The uses and impact of these tools could vary widely within an organization – and even within a given role. The third set of uses were in the background: generative AI features could be integrated into the background of various software systems, like Salesforce, that organizations already used, and the organization and its workers had not choice whether or not to "adopt" the AI tools.

At the organization level, we were interested in why some applications of generative AI scaled up and seemed to stick in the organization, while others were shelved. This was a harder question to ask at the individual level, where the patterns of usage in AI were hard to track. But from the perspectives of organizational leaders shaping AI adoption, there were several factors that seemed associated with scale-up.

The first common feature was that scaled-up uses of generative AI often addressed a pre-existing challenge that was already determined to be valuable to the organization. In healthcare, two cases of scaled-up applications are the tool that helped transcribe and annotate patient notes, as well as a tool that helped summarize patient records for handovers between shifts for nurses.

Both applications were responding to well-documented challenges within the healthcare sector: the outsized time that doctors were spending on medical notes, as well as the information loss that was occurring with handoffs between shifts. Similarly, the retailer application of generative AI to support in-store personnel with the knowledge that they need was a long-standing problem that previous technology was unable to solve. One factor that helps explain why these applications stick is that the workforce affected may already be bought in to the necessity of the tool. If frontline users have a shared interest and incentive to solve a problem, that might overcome their resistance to technological change.

A second common feature is that the solution to the organizational problem did not come from generative AI alone. There were often a constellation of complementary technologies – along with a human in the loop – that made the application work. For example, the technology company Zeta developed a customer service solution integrated LLMs to summarize incoming inquiries alongside robotic process automation (RPA) "bots" that they had developed to provide consistent responses to customer with different questions. The combination of the two technologies capitalized on the comparative advantages of each. The LLM-based tool could summarize and categorize the natural language inquiry; however, the company did not want the LLM to compose a response to each customer since they could not predict exactly what the LLM would say. The RPA bot that they developed would provide an identical response to each customer that asked a certain category of question. The LLM-based tool provided flexibility, and RPA provided reliability.

A third feature associated with scale-up is persistence. At BMS, the large life sciences company, the early stages of their application to use generative AI to speed up drug approval applications were unsuccessful. They were ready to shelve the application before company established an accelerator program to support teams across the organization to improve their AI applications incrementally over two-week periods. The program provided the promoted iterative experimentation with the AI tool until they were ready for production.

The three features make sense in tandem. Organizations that have a well-defined problem that they are committed to solving are more likely to find the right combination of technologies that can address the problem, and they are likely to develop support for teams of people to persist in finding an approach that works, even if the first applications do not provide promising results.

WHAT ABOUT AGENTS?

Near the end of our research, the organizations we studied began discussing the development of AI agents. The idea behind the shift was simple: rather than having individual workers ask questions of an AI tool and receive responses that can prompt the workers to act, the idea behind agents is that workers can prompt an AI tool to take an action, or series of actions. Workers themselves can design the AI tool to complete tasks. In healthcare, an agent may automatically follow steps to reconcile an insurance claim once it is filed in a system. In finance, an agent may automatically quote a price to insure an individual based on the application they submitted.

The promise of "agentic AI" has sometimes been presented as taking the human out of the loop, and unlocking more dramatic changes than the types of AI applications that we studied in the first years of experimentation. While we see early evidence that organizations are deploying agents to deepen their AI usage in areas where AI usage was already high – software development, customer service, administrative processes – agents do not represent a fundamentally new technological transformation for the organizations that we studied.

The agentic examples of AI that we have observed represent an incremental change from robotic process automation technology, where "bots" were programmed to take on routine tasks in a digital environment (often a series of clicks). When the bot could not handle the task – either because it was an edge case or something went wrong – the human reviewing the bot's workflow would solve the problem.

There are many routine tasks across the industries we studied where RPA technologies could automate a significant share of office work, turning human workers into supervisors of automated systems. However, the adoption of RPA technology has been slow, partly because the underlying routine processes often change, and the challenge of designing the right bot requires difficult coordination between those with technical expertise and those with expertise of complex business processes.

Advocates for agentic AI argue that RPA was limited because it could only perform routine tasks where the series of steps in a task could be codified in advance. Agentic AI can navigate non-routine and near-routine tasks based on an objective. This technical difference expands the tasks that agentic AI can potentially perform, but it also asks more of the human workers supervising AI agents. For most high-stakes tasks where agents encounter multiple possible ways of reaching an outcome – and where quality is important – the human in the loop overseeing agents ends up playing a critical role.

We already have examples of people in these roles overseeing automated systems with high variation, including operators in utility control rooms and mining remote operating centers, where automated systems and human oversight are necessary to ensure that the system is functioning well – and to anticipate potential challenges. Agentic AI represents a possibility that such human supervisory control roles will extend into new domains. But these technologies do not require new roles for humans in the loop altogether – nor do the examples that we have seen promise an elimination of humans in the loop for most tasks.

ANALYSIS

Our study of companies' uses of generative AI provides initial evidence on the three hypotheses suggesting how this set of technologies might be different.

Whereas there continue to be predictions that generative AI adoption will have dramatic short-term displacement effects on certain segments of workers, we did not see evidence of this pattern in the companies we studied. Across the companies we studied, applications of generative AI were deployed primarily with a "human in the loop" and not associated with significant layoffs or changes to the structure of work within the firm or industry. In most cases, the adoption of these technologies, although faster than technologies of the past, still represented incremental change in the first three years (of course, this says nothing about what may happen in the next three years).

There were several domains where generative AI use has seen dramatic short-term adoption, including customer service, software engineering, and marketing and communication. In these fields, our interviews suggest, the adoption of generative AI technologies has changed the nature of jobs and the expectations for workers. In software engineering, generative AI tools have been associated with a slowdown in hiring of junior workers, although it is unclear whether the deceleration of hiring is due to actual effects of the technology – or anticipated effects.

It remains unclear whether even these domains where AI has seen the most impact are supportive of the short-term displacement hypothesis. In customer service, for example, there is widespread use of generative AI tools, and only a small change in employment. Proponents of the hypothesis might point to declining job postings for customer service positions, similar to the relatively scarce job postings for entry-level software roles. However, the decline in customer service jobs (similar to the deceleration of software jobs) is oddly timed to the introduction of generative AI tools. For example, the decline of customer service job postings began a full year before ChatGPT was released.

Our observation that change to the workforce has been incremental is of course informed by the types of organizations that we studied – established firms, many in legacy industries, with large workforces that take time to transform. Even if these technologies were to have a dramatic long-term impact on the workforce in these industries, it would be unlikely that we would observe it in legacy companies in three years. It is possible that a transformative change will not happen by these companies creating new roles in some domains and shrinking roles in others – but by new companies emerging in these industries to compete with a completely different work organization.

Even with the incremental changes that we observed, there was a discernable change in the types of jobs that would be required if the organizations' applications of generative AI were successful. These new jobs would ask workers to take on a greater share of supervisory control tasks, overseeing automated systems, and fewer tasks where they manually operate processes. These types of jobs already exist in automated environments. For example, pilots perform supervisory control tasks in cockpits; air traffic controllers and utility operators in control rooms; and technicians on factory floors all handle some combination of these tasks. However, these tasks are relatively new in fields like law and medicine.

For the skill leveling hypothesis, the evidence from early studies of generative AI is clear. suggests those with less skill and experience before performing a task with generative AI appear to experience higher performance gains when they use generative AI. However, in our study, the impact of generative AI tools by skill level depends on the problem that the generative AI tool is set to address.

Applications that are designed to address the learning curve problem seem to be skill leveling. These tools are designed to extend the capabilities of novices in a given domain – not deepen expertise – so it is reasonable that the tools provide greater benefits the less skill the user has.

For the cafeteria and bottleneck problems, however, the skill effects appear different. Users of tools designed to solve these problems must interpret whether the output that the tool provides is relevant and accurate based on their pre-existing knowledge. In both cases, certain applications of AI may only be relevant to experienced workers in these domains.

Consider the BMS example, where workers who assemble compliance documents for pharmaceutical approvals must integrate expertise from across many documents and areas of expertise. A novice could do this job only if they relied on the support of human experts that they could trust. They could not rely on an AI tool because they would be unable to validate its reliability. A more experienced person may be able to use the generative AI tool because they could have the knowledge to validate and adjust the content that the tool provides.

There is a parallel issue with the bottleneck problem. In the medical example where doctors review patient notes automatically generated from a recording, the relevant skill for the user is determining whether the note is accurate and includes the appropriate annotations. Doctors with more experience should be able to review these results quickly because they know what a proper note should resemble.

There is similarly mixed evidence on the skill requirements hypothesis. Many of the organizations we studied have begun to assume that AI skills will become necessary to perform work across domains. However, it is not clear what those skills might be. In the current stage, the expectation is focused on outcomes: organizations aim for their workers to use AI tools to improve their performance at their jobs. The extent to which skill at prompting AI models or understanding how AI works "under the hood" contributes to success at using AI has yet to be determined. Part of the puzzle is that many new AI tools are so simple and intuitive to use that it is plausible that minimal additional training is necessary.

We did not find any strong evidence that organizations can use AI tools to leapfrog over gaps in their technical knowledge and infrastructure. Some organizations with messy internal data and documentation without a common format or database architecture have tried to use AI tools to extract information from their data universe. However, we did not find evidence of success for these experiments.

One challenge is that without a structured dataset that can be searched with traditional, deterministic data science tools, there is no clear reference point to ensure the accuracy of – and build trust in – the AI tool's results. Early research is consistent with this finding. An experiment with BCG consultants suggests that early AI tools may struggle to draw insights from a combination of data sources, including tabular and written data.

4. LEVERS FOR SHAPING GENERATIVE AI

There are three primary ways that applications of new technologies can fail to improve worker's performance, or the quality of their work: disuse (not automating when adding new technology could be beneficial), misuse (deploying automation technologies with poor results), and overuse (the automation technologies work well, but they lead to unintended consequences). The applications of generative AI that we have studied are vulnerable to all three, but there are ways to guard against them.

Our findings from studying early real-world experiments with generative AI, coupled with historical data on technology and work, suggest ten practical lessons for employers, policymakers, and the workforce.

The first lessons focus on the principles that can guide effective use of these tools.

1. **Gather evidence before scaling up.** We found that the most effective use cases of generative AI were for use cases where the problem was already defined. If an organization had already studied the problem and recognized that solving it would create improvements for the organization – either by improving the lives of their workers or speeding up a critical process for the company – then they had a clear vision of what generative AI needed to do. They knew what evidence they needed to see if generative AI worked. And once they had that evidence, they could begin to scale.

   This is not the only kind of generative AI experiment, however. Some explorations of generative AI have begun using these tools for all manner of tasks because they are new and exciting – and may seem to lead to improvement – but the hard evidence has been lacking. Perhaps there is a rough sense that the use of generative AI saves time on a task, but what are the consequences for quality and re-work? What other tasks are created? What is the downstream impact on the value added of the team responsible for the task? In some cases the evidence on all these questions is gathered and analyzed. In others, there is support for adoption before there is clear evidence that the use of generative AI tools are better than the alternative. [1]

2. **One size does not fit all, or even some.** Early evidence suggests that individual workers use generative AI tools in vastly different ways – even workers who are performing the same job. The flexibility of generative AI tools allow individuals to customize what they want out of the technology. This can create headaches for employers, such as the difficulty of measuring impact, but it is potentially a great benefit for job quality: workers can use these tools how they want, when they want, and can shun using them when they do not trust them to be effective.

   An approach that embraces the heterogeneity of generative AI use is consistent with survey data on how workers feel about automation and what makes a good job: some workers want to champion new technologies and value problem-solving at work. Other workers value collaboration and may be more skeptical of automated systems. Some employers have tried to select for champions of new technology, but there can be benefits to having a workforce with diverse usage of new technology. If there is variation in how workers use generative AI tools, then the organization can gather more data on what uses prove effective – and under what conditions – without jumping into a pattern of overuse.

3. **Learn when to trust.** An individual's willingness to trust is an important predictor of how effectively they use new technologies like robots and automated systems. Trust is also associated with an individual's optimism about new technologies. An ongoing challenge of generative AI systems is that users interacting with them cannot calibrate trust -- this is a key predictor of using a technology well, but there are inherent barriers to trusting a black box technology like generative AI.

   For generative AI tools to make workers faster at tasks and to improve the quality of their work, the individual users of generative AI will need to know when to trust these tools. There is active debate over whether the current generative AI tools can be truly transparent and explainable without additional fine-tuning or other modifications. If the technologies themselves remain opaque, then it is incumbent on organizations to establish practices and systems that workers can use to calibrate trust in the generative AI tools they use, so they can be confident when using these tools is appropriate – and when they should not be relying on them.

Based on these first principles around how to use generative AI tools, our research identified several outcomes that anchor the most promising applications of generative AI. What makes these outcomes appealing is that they represent benefits for individual workers and their careers, as well as for the performance of firms.

4. **Minimize drudgery.** Although some workers in surveys report enjoying their routine tasks, there are many workers who would prefer problem solving and creative tasks to the mundane and routine. These tasks are at the heart of the bottleneck problem. Some organizations have ventured into applications of generative AI that aim to augment tasks that workers enjoy – and from which they derive value – setting up resistance from the

---

1 Of course this is not the first time that new tools have been adopted for uses where the evidence was lacking. One of the most prominent is the adoption of digital devices for reading and note-taking despite consistent evidence that reading and retention on paper leads to more comprehension and retention than on screens. There are other reasons to read on screens, but particularly for learning to read, the evidence does not support the shift.

workers affected. The more straightforward applications of generative AI can focus on the bottlenecks that both workers and employers can agree would be better if they were gone. Employees closest to these tasks may be the best at identifying them, but employers can help identify the best technology to make these tasks less of a burden.

5. **Promote learning.** There is early evidence that using generative AI can lead to forms of mental offloading, where a worker performs a task using generative AI, but does not retain the knowledge of the task that they performed. In short, they are doing, but not learning. There are potential long-term consequences of this behavior for workers and organizations. Learning at work can unlock better career opportunities for employees, and deliver higher productivity for employers. There is evidence that generative AI technologies can help employees learn by pinpointing where there are gaps in their knowledge and providing relevant information. However, ensuring that generative AI tools are used for learning and not for offloading will require guardrails that steer workers in the former direction.

6. **Preserve teamwork.** Generative AI applications focused on addressing the cafeteria problem aim to empower individuals to perform a task alone that may have previously required a team. Although these applications are still early, one employer expressed concern that there would be a tradeoff between individual performance and team collaboration. Even in the cases where generative AI tools may empower solo acts, it may be important to preserve teamwork for other reasons, including mentorship, collective learning, and trust building among people who might rely on one another for other tasks.

The question remains: given these first principles and the ways that organizations may aim to transform work, what investments can be useful for helping achieve these goals? For instance, how can organizations push workers to use generative AI tools for learning and not as a shortcut? To calibrate trust rather than develop over-reliance?

7. **Interface design:** As employers have grown to understand the capabilities of generative AI tools, they have tended to buy tools from vendors rather than build their own custom applications. Even when organizations use tools built for their industry, they can customize the interfaces that their workers use to access these tools.

   Interfaces have long been able to shape how workers interact with new technologies, steering user behavior and forming user understanding. The principles for good interface design have sought to maximize a user's situational awareness – their ability to understand what is happening and why – as well as to manage their mental workload. Organizations deploying generative AI can test how their workers respond to various interface designs by using existing measures like the Situational Awareness Global Assessment Technique (SAGAT), which can estimate retention of information, and the NASA TLX system to measure workers' mental workload.

8. **Domain expertise:** There is some indication that generative AI tools can help users extend their capabilities. However, this should not be mistaken as a replacement for domain expertise. The users of AI tools must still be able to provide the underlying context, interpret their results, and improve them as they improve their understanding of the background domain. In short, domain expertise can still complement generative AI technologies, but the expertise itself must be sufficient to validate and scrutinize what generative AI tools provide.

   The need for domain expertise to use these tools at their fullest potential leads to a perhaps counterintuitive conclusion: that training institutions should invest more in developing domain experts in areas with high potential for generative AI to make an impact like medicine, computer science, and the life sciences. In the short term, there may be lower demand for entry-level support in these industries. However, in the long-term, a potential bottleneck to breakthroughs in these domains is that generative AI tools can produce a high volume of interesting new techniques, technologies, and medicines, but only if there are domain experts to solicit, interpret and test these concepts. These fields may also require domain experts to navigate complex bureaucratic and regulatory processes that will require trust-building and interpersonal explanation. Realizing the high potential of generative AI tools might require more people – not fewer – in these areas.

9. **Accountability:** Generative AI tools can create a moral hazard for workers using them. A worker may have an incentive to use generative AI tools as a shortcut to produce information and perform work that appears valid and impressive, even if that work contains errors or masks a lack of underlying understanding or capability. Since many organizations currently operate on a trust relationship between those analyzing information and those interpreting it, it is unlikely that a worker using generative AI will have any of their analysis challenged – or errors uncovered and traced back to them. Nonetheless, the benefits of learning and the costs of error-riddled analyses are both high.

One way for organizations to address the moral hazard challenge is to establish accountability for workers using AI tools to increase their incentives for learning and raise their costs of making an error. One approach to this problem comes from airport security, where random additional screenings and can focus the worker's attention and validate that the information from the automated system is valid and complete.

10. **New work.** The most common fear with the introduction of new technology is that it will displace work. However, there is a less storied bright side of new technologies: the proliferation of many and varied new jobs. It is not always clear how these jobs are created: some appear to evolve naturally based on organizational needs, whereas others are hatched through an entrepreneurial process.

    When workers adopt generative AI tools that change the tasks they do at work – and how they do them – often freeing up part of their days, it is an opportunity for their employers to re-design the jobs that they offer. This should not just be about what tasks and skills organizations need in the moment. It should also be about responding to the tasks that workers find most enjoyable and the skills that they are most excited to develop. Designing new jobs around the needs of the organization – as well as the preferences of workers – can help cultivate the type of employer-employee relationship that leads to workers using technology effectively. When workers feel that their employer is investing in them – and they are motivated to learn and grow in their career – they are also more likely to be optimistic about the role that technology plays in their careers.