GenAI Doesn't Just Increase Productivity. It Expands Capabilities.

SEPTEMBER 05, 2024

By Daniel Sack, Lisa Krayer, Emma Wiles, Mohamed Abbadi, Urvi Awasthi, Ryan Kennedy, Cristián Arnolds, and François Candelon

READING TIME: 12 MIN

This is the second major field experiment led by the BCG Henderson Institute designed to help business leaders understand how humans and GenAI should collaborate in the workplace. Our previous study assessed the value created—and destroyed—by GenAI when used by workers for tasks they had the

capabilities to complete on their own. Our latest experiment tests how workers can use GenAI to complete tasks that are beyond their current capabilities.

A new type of knowledge worker is entering the global talent pool. This employee, augmented with generative AI, can write code faster, create personalized marketing content with a single prompt, and summarize hundreds of documents in seconds.

These are impressive productivity gains. But as the nature of many jobs and the skills required to do them evolve, workers will need to expand their current capabilities. Can GenAI be a solution there as well?

Based on the results of a new experiment conducted by the BCG Henderson Institute and scholars from Boston University and OpenAI's Economic Impacts research team, the answer is an unequivocal yes. We've now found that it's possible for employees who didn't have the full know-how to perform a particular task yesterday to use GenAI to complete the same task today.

METHODOLOGY ─

Our research involved a carefully structured experimental design to evaluate the impact of generative AI on the ability of nontechnical knowledge workers to perform technical, data-science tasks.

In total, 480 BCG consultants and 44 BCG data-scientist volunteers completed this controlled study. The study participants were general consultants, for whom data-science expertise is not typically required. This expertise exists among the data scientists of BCG X who also volunteered to support the study by establishing benchmarks. The performance of general consultants was evaluated by comparing their output to that of BCG data scientists who completed the same tasks.

General consultants were randomly assigned to either a GenAI-augmented group, which received interactive training on using Enterprise ChatGPT-4 with the Advanced Data Analysis Feature for data science tasks, or a control group, which was asked not to use GenAI and received interactive training on traditional resources like Stack Overflow. The tasks assigned to participants included coding, statistical understanding, and predictive modeling, all of which required skills that are typically outside the expertise of nontechnical workers but within the day-to-day expertise of BCG's data scientists.

Data collection was carried out in four phases: a pre-experiment survey to assess baseline skills and attitudes, a tailored training session for each group, the completion of two out of three randomly assigned data-science tasks, and a post-experiment survey to measure knowledge retention without the use of AI tools.

The tasks were designed by BCG data scientists to ensure that they were challenging enough that AI could not solve them independently. Analysis focused on comparing the performance of the treatment and control groups against the benchmarks set by BCG data scientists, examining the completion rates, time taken, and correctness of the responses.

Similar to our first GenAI study, we "put our feet to the fire"—with a goal of deeply understanding GenAI's impact on ways of working.

With that in mind, leaders should embrace GenAI not only as a tool for increasing productivity, but as a technology that equips the workforce to meet the changing job demands of today, tomorrow, and beyond. They should consider generative AI an exoskeleton: a tool that empowers workers to perform better, and do more, than either the human or GenAI can on their own.
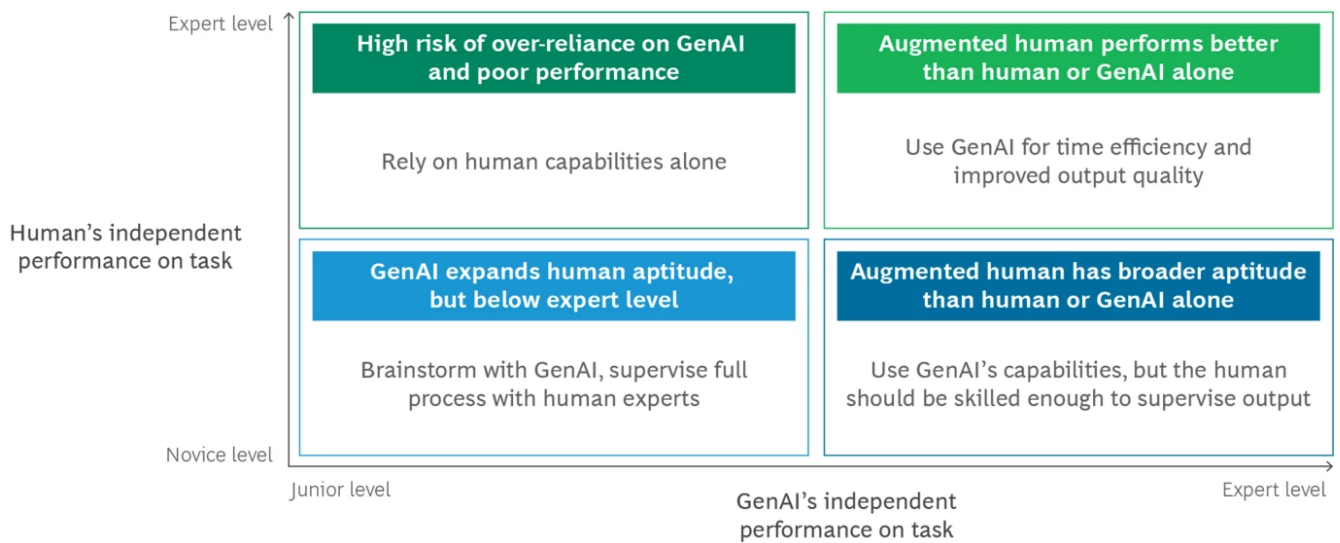
Of course, there are important caveats—for example, employees may not have the requisite knowledge to check their work, and therefore may not know when the tool has gotten it wrong. Or they may become less attentive in situations where they should be more discriminating.

But leaders who effectively manage the risks can reap significant rewards. The ability to rapidly take on new types of work with GenAI—particularly tasks that traditionally require niche skills that are harder to find, such as data science—can be a game-changer for individuals and companies alike.

How GenAI Can Equip Knowledge Workers

In the previous experiment, we measured performance on tasks that were within the realm of the participants' capabilities.[1] (See top row of Exhibit 1.) For tasks where GenAI is highly capable, we found that augmented workers perform significantly better than humans working without the technology. However, when the technology is not capable of performing the task at expert level, humans tend to over-rely on GenAI and perform worse than if they had completed the task on their own.

Exhibit 1 - When and How to Pair Humans and GenAI

Expert level

High risk of over-reliance on GenAI and poor performance	Augmented human performs better than human or GenAI alone
Rely on human capabilities alone	Use GenAI for time efficiency and improved output quality

Human's independent performance on task

GenAI expands human aptitude, but below expert level	Augmented human has broader aptitude than human or GenAI alone
Brainstorm with GenAI, supervise full process with human experts	Use GenAI's capabilities, but the human should be skilled enough to supervise output

Novice level

Junior level ——————————————————→ Expert level

GenAI's independent performance on task

Source: BHI analysis.

But what happens when, instead of using GenAI to improve performance within their current skillset, people use GenAI to complete tasks that are outside their own capabilities? Does being augmented with GenAI expand the breadth of tasks people can perform?

For our latest experiment, more than 480 BCG consultants performed three short tasks that mimic a common data-science pipeline: writing Python code to merge and clean two data sets; building a predictive model for sports investing using analytics best practices (e.g. machine learning); and validating and correcting statistical analysis outputs generated by ChatGPT and applying statistical metrics to determine if reported findings were meaningful.[2]

While these tasks don't capture the entirety of advanced data scientists' workload, they are sufficiently representative. They were designed to present a significant challenge for any consultant and could not be fully automated by the GenAI tool.[3]
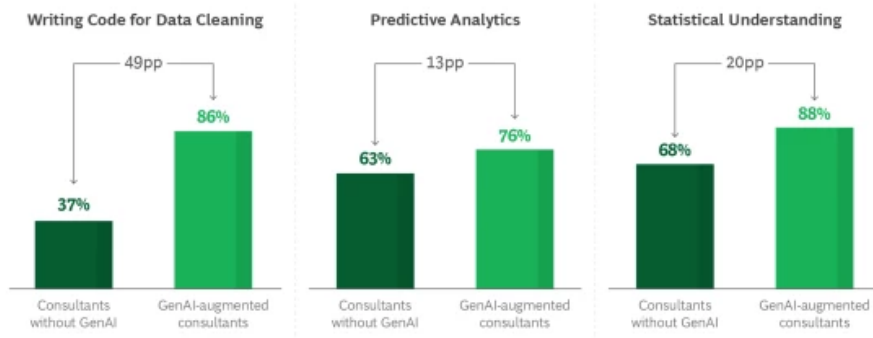
To help evaluate the performance impact of GenAI, only half of the participants were given access to the GenAI tool, and we compared their results to those of 44 data scientists who worked without the assistance of GenAI. When we dive deeper into the results, three critical findings emerge.

The Immediate Aptitude-Expansion Effect

When using GenAI, the consultants in our study were able to instantly expand their aptitude for new tasks. Even when they had no experience in coding or statistics, consultants with access to GenAI were able to write code, appropriately apply machine learning models, and correct erroneous statistical processes.[4] (See Exhibit 2.)

Exhibit 2 - GenAI Significantly Improved Performance in Three Data-Science Tasks

Performance of consultants on tasks outside their capabilities[1]

Writing Code for Data Cleaning	Predictive Analytics	Statistical Understanding

Writing Code for Data Cleaning
- 49pp
- Consultants without GenAI: 37%
- GenAI-augmented consultants: 86%

Predictive Analytics
- 13pp
- Consultants without GenAI: 63%
- GenAI-augmented consultants: 76%

Statistical Understanding
- 20pp
- Consultants without GenAI: 68%
- GenAI-augmented consultants: 88%

Sources: Boston University; OpenAI's Economic Impacts research team; BHI analysis.

[1] All scores are normalized so that 100% is equivalent to the benchmark set by the average scores of participating data scientists.

We observed the biggest aptitude-expansion effect for coding, a task at which GenAI is highly adept. Participants were asked to write code that would clean two sales data sets by correcting missing or invalid data points, merging the data sets, and filtering to identify the top five customers in a specified month.

Participants who used GenAI achieved an average score equivalent to 86% of the benchmark set by data scientists. This is a 49-percentage-point improvement over participants not using GenAI. The GenAI-augmented group also finished the task roughly 10% faster than the data scientists.

Even those consultants who had never written code before reached 84% of the data scientists' benchmark when using GenAI. One participant who had no coding experience told us: "I feel that I've become a coder now and I don't know how to code! Yet, I can reach an outcome that I wouldn't have been able to otherwise." Those working without GenAI, on the other hand, often did not get much further than opening the files and cleaning up the first "messy" data fields; they achieved just 29% of the data-scientist benchmark.

It's important to note that most consultants are expected to know the basics of data cleaning and often perform data-cleaning tasks using no-code tools such as Alteryx. Therefore, while they did not have experience doing the coding task in Python, they knew what to expect from a correct output. This is critical for any GenAI-augmented worker—if they don't have enough knowledge to supervise the output of the tool, they will not know when it is making obvious errors.

A Powerful Brainstorming Partner

For the task that involved predictive analytics, our participants faced a challenging scenario: neither they nor the GenAI tool were highly adept at that task. Here, the technology was still valuable as a brainstorming partner.

While all the tasks in our experiment were designed such that the GenAI could not independently solve them, the predictive-analytics task required the most engagement from participants. They were asked to create a predictive model, using historical data on international soccer matches, to develop an investment strategy. Their ultimate goal was to assess how predictable, or reliable, their model would be for making investment decisions.

> **Many participants used GenAI to brainstorm, combining their knowledge with the tool's knowledge to discover new modeling and problem-solving techniques.**

As shown in Exhibit 2, this was the task on which the GenAI-augmented consultant was least likely to perform on par with a data scientist, regardless of previous experience in coding or statistics. This is because the GenAI tool is likely to misunderstand the ultimate goal of the prompt if the entire task is copied and pasted directly into the tool without breaking the question into parts or clarifying the goals. As a result, participants with access to GenAI were more likely to be led astray than their nonaugmented counterparts.

Even so, we found that, with the support of GenAI, many participants were able to step outside their comfort zone. They brainstormed with the tool, combining their knowledge with GenAI's knowledge to discover new modeling techniques and identify the correct steps to solve the problem successfully. The GenAI-augmented participants were 15 percentage points more likely to select and appropriately apply machine-learning methods than their counterparts who did not have access to GenAI.

Reskilled, but Only When Augmented

Participants' aptitude for completing new and challenging tasks was immediately boosted when using GenAI, but were they reskilled? Reskilling is defined as an individual gaining new capabilities or knowledge that enables him or her to move into a new job or industry. We found in our study that GenAI-augmented workers were in a sense "reskilled," in that they gained new capabilities that were beyond what either the human or GenAI could do on their own. But GenAI was only an exoskeleton; the participants were not intrinsically reskilled, because "doing" with GenAI does not immediately nor inherently mean "learning to do."

While each participant was assigned just two of the three tasks in the experiment, we gave everyone a final assessment with questions related to all three tasks to test how much they actually learned. For example, we asked a coding syntax question even though not everyone did the coding task—and therefore not everyone would have had a chance to "learn" syntax. Yet the people who participated

in the coding task scored the same on the assessment as people who didn't do the coding task. Performing the data-science tasks in our experiment thus did not increase participants' knowledge.

Of course, participants only had 90 minutes to complete the task. With repetition, more learning might have occurred. We also didn't inform participants that they would be tested at the end, so incentivizing learning might also have helped. This is important, because we found that having at least some background knowledge of a given subject matters.

> **We found that coding experience is a key success factor for workers who use GenAI—even for tasks that don't involve coding.**

GenAI-augmented participants with moderate coding experience performed 10 to 20 percentage points better on all three tasks than their peers who self-identified as novices, *even when coding was not* involved.[5] In fact, those with moderate coding experience were fully on par with data scientists for two of the three tasks—one of which had zero coding involved.
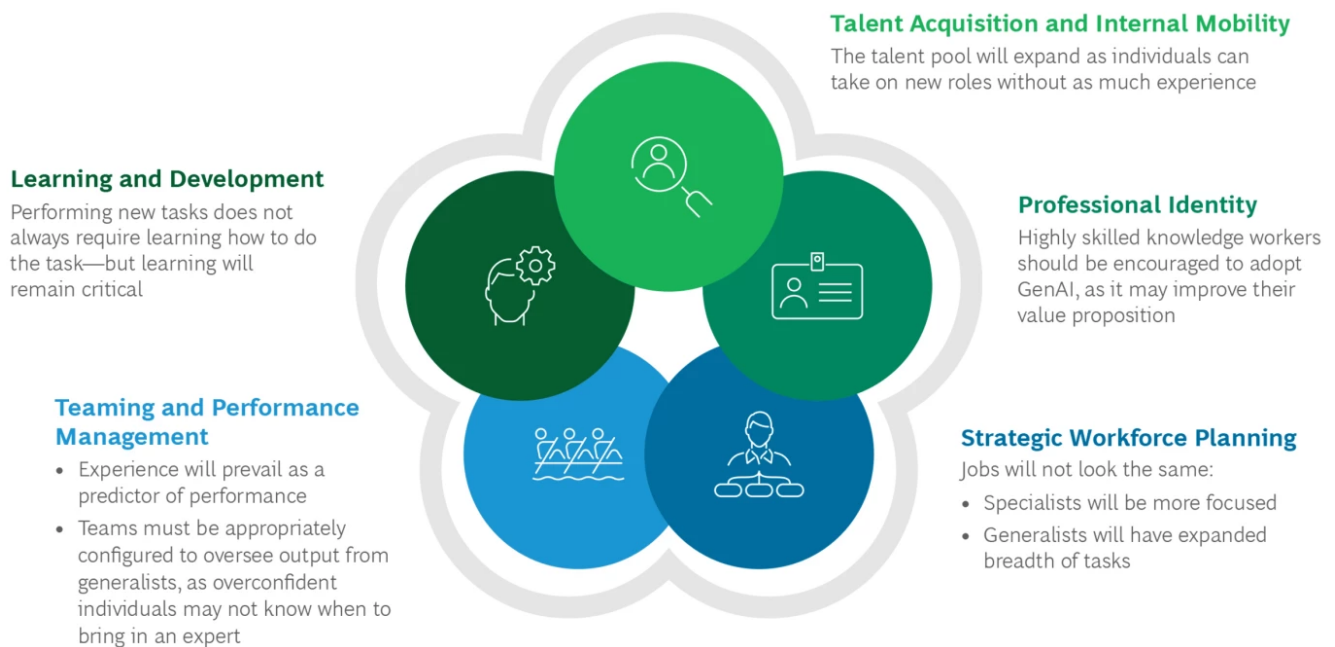
Based on this, we posit that it is the engineering mindset that coding helps develop—for example, having the ability to break a problem down into subcomponents that can be effectively checked and corrected—that ultimately matters, more so than the coding experience itself.

The risk of fully automating code, then, is that people don't form this mindset—because how do you maintain this skill when the source of its development is no longer needed? This is part of a larger discussion: What other seemingly automatable skills have such importance? Will these skills become the new Latin, taught mostly to cultivate a particular mindset?

Managing the Transition

While we have used data science as a case study, we believe that our finding—that augmented workers can skillfully perform new tasks—can be applied to any field that is within the tool's capabilities. We've identified five core implications for company leaders. (See Exhibit 3.)

Exhibit 3 - Workforce and Change-Management Implications

Talent Acquisition and Internal Mobility
The talent pool will expand as individuals can take on new roles without as much experience

Learning and Development
Performing new tasks does not always require learning how to do the task—but learning will remain critical

Professional Identity
Highly skilled knowledge workers should be encouraged to adopt GenAI, as it may improve their value proposition

Teaming and Performance Management
- Experience will prevail as a predictor of performance
- Teams must be appropriately configured to oversee output from generalists, as overconfident individuals may not know when to bring in an expert

Strategic Workforce Planning
Jobs will not look the same:
- Specialists will be more focused
- Generalists will have expanded breadth of tasks

Source: BHI analysis.

Talent Acquisition and Internal Mobility. The results across our workforce experiments have shown that what an individual can perform on his or her own by no means approaches what can be accomplished when augmented by GenAI. This suggests that the talent pool for skilled knowledge work is expanding.

Recruiters should therefore incorporate GenAI into the interview process to get a more complete picture of what a prospective employee might be capable of when augmented by the technology. Leaders may also find that an unlikely person inside their organization can fill an open role. We're not suggesting that nontechnical generalists can immediately become data scientists. But a generalist marketer could, for example, take on marketing analyst tasks or roles.

Learning and Development. What does this mean for employees seeking paths to senior roles and/or leadership? How should members of the GenAI-augmented workforce, who can flexibly take on various roles, cultivate the right skills for career advancement—and what are the most important skills for them to retain long term?

While GenAI has an immediate aptitude-expansion effect, learning and development remain the most import lever for cultivating advanced skills and supporting each employee's professional trajectory. Leaders therefore must ensure that employees have incentivized and protected time to learn. Other research has shown that when specifically used for learning (and, unlike our participants, people are generally incentivized to learn in their jobs), GenAI is an effective personalized training tool.

> **Leaders should ensure that future implementations of GenAI tools include the functionality to inform the user if a task is outside the technology's capability set.**

Our analysis also suggests that developing some technical skills leads to greater performance, even for nontechnical workers. Regardless of the training employees receive, company leaders should ensure their future implementations of GenAI tools include the functionality to inform the user if a task is outside the technology's capability set—information that should be compiled from regular benchmarking.

Companies are likely to find competitive advantage from developing tools and processes that precisely assess the capabilities of GenAI models for their use cases. As shown in Exhibit 1, how a worker should use GenAI greatly depends on understanding where a task lies within their own skill set and within the capabilities of the technology.

Teaming and Performance Management. Although our results show it is possible for a generalist to take on more complex knowledge work, it will be crucial to manage their performance and ensure the quality of their output. This could mean designing cross-functional teams to provide generalists with easy access to an expert when they need help and establishing regular output-review checkpoints—because an overconfident generalist may not always know when to ask for support. Leaders will need to run pilots to ensure their teaming configurations lead to the best outcomes. This may be an opportunity to break silos and integrate teams of generalists with experts from various centers of excellence.

Strategic Workforce Planning. Given the implications for talent and teaming, how should organizations think about specialized expert tracks and the structure of their workforce? What does strategic workforce planning for knowledge work mean in a world of constant job transformation and technological advancement? We don't have all the answers. But we do see that the skills needed for a given role are blurring, and workforce planning will no longer be solely focused on finding a certain number of people with a specific knowledge skill, such as coding.

Instead, planning should include a focus on behavioral skills and enablers that will support a more flexible workforce. While knowledge workers may be technically capable of taking on new roles with the help of GenAI, not everyone is equally adept at embracing change.

Professional Identity. The impact of GenAI on professional identity is an important and contentious topic. But a recent survey suggests that negative impacts can be mitigated when employees feel supported by their employers.

In fact, in our study, we found that 82% of consultants who regularly use GenAI for work agree with the statements "Generative AI helps me feel confident in my role" and "I think my coworkers enjoy using GenAI for their work," compared to 67% of workers who don't use it on a weekly basis. More than 80% of participants agreed that GenAI enhances their problem-solving skills and helps them achieve faster outputs.

This suggests that highly skilled knowledge workers genuinely enjoy using the tool when it allows them to feel more confident in their role—which aligns with our previous findings that mandating the use of AI can actually improve employee perception of AI. However, this is *only* true if employees believe that AI is being deployed to their benefit.

We are only at the beginning of the GenAI transformation journey, and the technology's capabilities will continue to expand. Executives need to be thinking critically about how to plan for this future, including how to redefine expertise and what skills to retain in the long term.

But they are not alone: Skill development is a collaborative effort that includes education systems, corporate efforts, and enablement platforms such as Udemy and Coursera. Even the providers of GenAI models should be thinking about how their tools can further enable learning and development. Preparing for the GenAI-augmented workforce must be a collective endeavor— because our collective future depends on it.

BCG **HENDERSON INSTITUTE**

The BCG Henderson Institute is Boston Consulting Group's strategy think tank, dedicated to exploring and developing valuable new insights from business, technology, and science by embracing the powerful technology of ideas. The Institute engages leaders in provocative discussion and experimentation to expand the boundaries of business theory and practice and to translate innovative ideas from within and beyond business. For more ideas and inspiration from the Institute, please visit our website and follow us on LinkedIn and X (formerly Twitter).

Authors
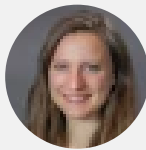
Daniel Sack

MANAGING DIRECTOR & PARTNER

Stockholm

Lisa Krayer

PRINCIPAL

Washington, DC

Emma Wiles

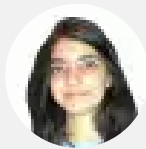**ASSISTANT PROFESSOR OF INFORMATION SYSTEMS, BOSTON UNIVERSITY'S QUESTROM SCHOOL OF BUSINESS**
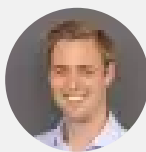
Mohamed Abbadi

CONSULTANT

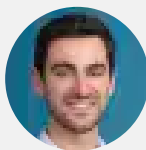Washington, DC

Urvi Awasthi

DATA SCIENTIST

New York

Ryan Kennedy

AI ENGINEER

Boston
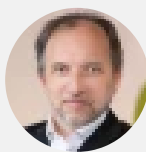
Cristián Arnolds

CONSULTANT

New York

François Candelon

ALUMNUS

1 That experiment was conducted using the first version of GPT-4.

2 Of the consultants who originally signed up to participate, 480 completed the experiment. Participants were randomly split into a control group that was not allowed to use GenAI for the tasks and a "treatment" group that was asked to use GenAI. Each participant was randomly assigned two of the three tasks; each task was timeboxed for 90 minutes. It is important to note that we did not test a full end-to-end data science workflow from ideation to delivery or advanced topics such as deep learning.

3 We used Enterprise ChatGPT with GPT-4 and its Advanced Data Analysis feature.

4 The findings for the statistical understanding task were consistent with those of the coding and predictive analytics tasks; as a result, we focused on the first two tasks in this article.

5 Coding experience was based on a self-assessment. We define moderate experience by those who selected "I know how to code but am not an expert" and novices as those who selected "I only know the basics of coding" and "I don't know how to code."

ABOUT BOSTON CONSULTING GROUP

Boston Consulting Group partners with leaders in business and society to tackle their most important challenges and capture their greatest opportunities. BCG was the pioneer in business strategy when it was founded in 1963. Today, we work closely with clients to embrace a transformational approach aimed at benefiting all stakeholders—empowering organizations to grow, build sustainable competitive advantage, and drive positive societal impact.

Our diverse, global teams bring deep industry and functional expertise and a range of perspectives that question the status quo and spark change. BCG delivers solutions through leading-edge management consulting, technology and design, and corporate and digital ventures. We work in a uniquely collaborative model across the firm and throughout all levels of the client organization, fueled by the goal of helping our clients thrive and enabling them to make the world a better place.

For information or permission to reprint, please contact BCG at permissions@bcg.com. To find the latest BCG content and register to receive e-alerts on this topic or others, please visit bcg.com. Follow Boston Consulting Group on Facebook and X (formerly Twitter).