



Organization Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork

Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz, Raffaella Sadun, Ethan Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, Karim R. Lakhani

To cite this article:

Fabrizio Dell'Acqua, Charles Ayoubi, Hila Lifshitz, Raffaella Sadun, Ethan Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, Karim R. Lakhani (2026) The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork. *Organization Science*

Published online in Articles in Advance 12 Jun 2026

<https://doi.org/10.1287/orsc.2025.20702>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Organization Science*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/orsc.2025.20702>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2026 The Author(s)

Please scroll down for article—it is on subsequent pages









With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork

Fabrizio Dell'Acqua,^{a,b,*} Charles Ayoubi,^c Hila Lifshitz,^{b,d} Raffaella Sadun,^{a,b} Ethan Mollick,^e Lilach Mollick,^e Yi Han,^f Jeff Goldman,^f Hari Nair,^f Stew Taub,^f Karim R. Lakhani^{a,b}

^aHarvard Business School, Harvard University, Boston, Massachusetts 02134; ^bHBS AI Institute, Harvard University, Boston, Massachusetts 02163; ^cESSEC Business School, 95000 Cergy, France; ^dWarwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom; ^eThe Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; ^fProcter & Gamble, Cincinnati, Ohio 45202

*Corresponding author

Contact: fdellacqua@hbs.edu,  <https://orcid.org/0000-0002-1998-0542> (FDA); ayoubi@essec.edu,  <https://orcid.org/0000-0002-0752-3328> (CA); hdiginnovation@gmail.com,  <https://orcid.org/0000-0002-3461-003X> (HL); rsadun@hbs.edu,  <https://orcid.org/0000-0002-0987-8923> (RS); emollick@wharton.upenn.edu,  <https://orcid.org/0000-0001-6231-496X> (EM); lmollick@wharton.upenn.edu (LM); han.y.7@pg.com (YH); goldman.js@pg.com (JG); nair.h.1@pg.com (HN); taub.sl@pg.com (ST); klakhani@hbs.edu,  <https://orcid.org/0000-0002-5535-8304> (KRL)

Received: June 3, 2025

Revised: November 17, 2025; March 26, 2026


Accepted: April 1, 2026

Published Online in *Articles in Advance*:
June 12, 2026

<https://doi.org/10.1287/orsc.2025.20702>

Copyright: © 2026 The Author(s)

Abstract. We examine how artificial intelligence (AI) impacts three core pillars of collaboration—performance enhancement, expertise integration, and social engagement—through a preregistered field experiment with 791 professionals at Procter & Gamble, a global consumer packaged goods company. Working on real product innovation challenges, professionals were randomly assigned to work either with or without AI, and either individually or with another professional in new product development teams. Our findings show that (1) AI significantly enhances performance: individuals with AI matched the performance of teams without AI, suggesting that AI can effectively replicate certain benefits of human collaboration. Moreover, (2) AI helps bridge functional silos: without AI, research and development professionals tended to suggest more technical solutions, whereas commercial professionals leaned toward commercially oriented proposals. Professionals using AI produced more balanced solutions, regardless of their professional background. (3) AI's language-based interface prompted more positive self-reported emotional responses among participants, suggesting it can fulfill part of the social and motivational role traditionally offered by human teammates. Finally, decomposing the innovation process suggests that AI primarily enhances the quality of generated ideas, shifting the distribution of creative output upward, whereas human judgment retains value in evaluative selection. This finding highlights the multiple and complementary roles that human and AI partners can play in new product development tasks and creative problem solving. More generally, our results suggest that AI adoption in knowledge work affects not only performance but also how expertise and sociality appear within teams, offering insights into the impact of generative AI on collaborative work within organizations.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Organization Science*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/orsc.2025.20702>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Funding: Funding for this research was provided in part by Harvard Business School.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/orsc.2025.20702>.

Keywords: technology and innovation management • research design and methods • field experiments • implementation of new technology • organization and management theory • organizational processes • economics and organization • organizational economics

1. Introduction

Teamwork is central to modern organizations. Whether designing a new product, solving strategic challenges, or supporting impactful innovation, human collaboration has often been associated with higher-quality results than individuals working alone (Wuchty et al. 2007, Singh and Fleming 2010). There are three main pillars commonly used to justify teamwork. The first is

performance: teamwork is often more effective than individual work and allows for more complex problems to be tackled (Ancona and Caldwell 1992, Lindbeck and Snower 2000, Deming 2017, Weidmann and Deming 2020). The second is expertise sharing and knowledge complementarities: teamwork allows people with different expertise to come together and work on the same problem effectively (Kogut and Zander 1992,

Argote 1999, Faraj and Sproull 2000, Nickerson and Zenger 2004). Finally, there is human sociality: people often enjoy connecting with other people, which can increase their motivation to work (Deutsch 1949, Johnson and Johnson 2005, Kozlowski and Bell 2013). In this paper, we investigate how the introduction of artificial intelligence (AI) impacts these three pillars of teamwork in the context of a field experiment in a large organization.

The integration of AI into knowledge work poses an important challenge: although AI, particularly generative AI (GenAI), has demonstrated the capacity to enhance individual creativity, productivity, and decision making (Dell'Acqua et al. 2023, Noy and Zhang 2023, Peng et al. 2023, Brynjolfsson et al. 2025), its ramifications for team-based collaboration remain largely unexplored. Prior work has treated AI primarily as a tool, like a spreadsheet or calculator, that can be used to enhance individual performance. But a unique aspect of large language models (LLMs), the most common form of GenAI, is that they are trained on human language and often act more like a person than a machine (Mollick 2024). This leads to a key question: Can GenAI fill some of the roles of humans in teamwork in specific collaborative contexts? We examine this by moving past considering AI as a mere tool, but instead ask whether it can provide some of the same benefits of human teamwork, namely, collective performance, expertise sharing, and social connection.

Understanding these dimensions has profound implications for organizational learning and innovation strategies. As companies integrate GenAI into day-to-day work (Bick et al. 2026), its influence may extend beyond one-off productivity gains: it may affect how knowledge is shared and recombined across functional lines and how knowledge workers experience their work. By examining AI as a potential collaborator rather than simply a tool, our research considers how these technologies might affect knowledge work by facilitating knowledge diffusion, reducing functional silos, and creating more positive emotional experiences during innovation activities—all factors that may matter for organizational learning.

To do so, we designed a field experiment exploring three main dimensions. (1) Does GenAI provide the performance gains traditionally attributed to teamwork? (2) Does GenAI enable a broadening of expertise even when employees lack some specialized knowledge? Finally, (3) can GenAI offer the kind of social engagement that we typically associate with human collaboration? Put simply, to what extent can AI be treated as a “cybernetic teammate,” rather than as yet another software tool?

Our research addresses these questions through a field experiment and organizational upskilling program involving 791 experienced professionals at Procter &

Gamble (P&G), a global consumer packaged goods company with roughly 7,000 research and development (R&D) professionals worldwide. The experiment was developed in close collaboration with the company over the course of a year, requiring sustained coordination with senior leadership to design a protocol that mirrored the company's actual new product development routines. R&D and commercial professionals dedicated a full working day to the study, engaging with real, long-standing business challenges from their own units, not hypothetical tasks. Participation carried real professional stakes: the best proposals would be presented to business unit leaders. The evaluation process was based on assessments by multiple independent evaluators, without time constraints, under a protocol validated by P&G managers. The best ideas generated were of sufficient quality to enter the company's actual innovation pipeline. The resulting data set is grounded in existing business problems, within the company's core organizational processes, and with strong professional incentives.

Our participants were randomly assigned to one of four conditions, in a 2×2 experimental design: (1) an individual working without GenAI, (2) a team of two humans without GenAI, (3) individuals with GenAI, and (4) a team of two humans plus GenAI. All teams comprised one commercial professional and one R&D professional, ensuring authentic cross-functional collaboration that reflects real-world organizational structures.¹ Each individual or team was assigned to develop a new solution to address a real business need for their business unit, ensuring they could leverage their domain expertise on the business needs they regularly target in their work. The build-up of teams working on early-stage specific product development problems resembles “flash teams” (Valentine and Bernstein 2025) and enables precise causal identification while remaining embedded within the organization.

Within this framework, we focus on three main outcomes that map onto the pillars of teamwork. First, we examine performance: Can AI help people produce high-quality work in innovative product development, potentially with better ideas or more thorough exploration of solutions? Second, we look at expertise: Does AI enable participants to breach typical functional boundaries—for instance, allowing R&D professionals to produce commercially viable ideas or commercial professionals to propose technically sound solutions? Third, we measure human sociality. Although this can take many forms, we operationalize it as the emotional dimensions of the collaborative experience. Specifically, we ask, To what extent does AI actually affect emotional experiences—such as excitement, engagement, or frustration—that traditionally emerge from human-to-human interaction?

Our findings show that AI reproduces some of the benefits of human collaboration, acting as a cybernetic teammate.² Individuals with AI produce solutions at a quality level comparable to two-person teams, indicating that AI can indeed stand in for certain collaborative functions. The adoption of AI also broadens the user's reach in areas outside their core expertise. Workers without deep product development experience, for example, can leverage AI's suggestions to bridge gaps in knowledge or domain understanding, reproducing some of the knowledge integration typically achieved through human collaboration. This has the potential to diminish functional boundaries, improving access to expertise within teams and organizations. Moreover, professionals reported more positive emotions and fewer negative emotions when engaging with AI compared with working alone, echoing some of the emotional benefits traditionally associated with human teamwork. This pattern differs from previous findings about technology's typically negative impact on workplace social dynamics.

We also examine how AI affects the innovation process itself. New product development involves multiple stages—generating ideas, selecting among them, and developing chosen concepts into detailed solutions. By analyzing performance at each stage, we find that AI's benefits stem primarily from enhancing idea generation quality rather than improving selection accuracy. AI shifts the distribution of creative output upward, producing better ideas from the outset while preserving variance in outcomes. This pattern suggests AI functions as a quality amplifier in collaborative work, improving the raw material of innovation while human evaluative judgment continues to play a role in identifying the most promising concepts. These findings may help motivate future research on more elaborate collaborations within organizations.

Overall, our findings indicate that adopting AI in knowledge work involves more than simply adding another tool. By enhancing performance, blurring functional boundaries, and altering emotional experiences, the introduction of GenAI into organizations may affect how organizations structure teams and define individual roles. As firms integrate AI technologies more widely, they must consider not only operational efficiencies but also emotional and social implications for workers. Our study provides a basis for understanding these shifts and offers insights that may help inform the design of AI-enhanced work environments, where AI can play an active collaborative role.

2. Theoretical Background

Organizations face a critical implementation challenge with the emergence of GenAI: how to effectively integrate this powerful new technology into

team workflows. Unlike previous technologies that primarily automated routine tasks, GenAI's capacity to converse, reason, and create positions it less as a passive tool and more as a cybernetic teammate whose presence alters team design, as it can be conceptualized as an active participant in collaborative processes. This perspective requires organizations to consider the complex dynamics of human–AI team integration, drawing on what we know about effective team functioning—including knowledge sharing and human sociality. Additionally, the introduction of GenAI as a teammate may significantly impact organizational learning trajectories, potentially restructuring how knowledge is created and retained within teams.

2.1. Teamwork

The nature of knowledge work is becoming ever more collaborative (Lazer and Katz 2003, Deming 2017, Puranam 2018). Teamwork forms the backbone of modern organizations for multiple reasons, but foremost among them is performance. A wide range of scholarship shows that collaboration can outperform individual effort in organizations by integrating multiple perspectives, thereby tackling complex problems more effectively (Ancona and Caldwell 1992, Cohen and Bailey 1997, Csaszar 2012). Although collaborative production creates unique organizational challenges (Alchian and Demsetz 1972), Cohen and Bailey (1997) highlight that well-structured teamwork can mobilize broad-based knowledge under high task complexity. In the same vein, Csaszar (2012) demonstrates how collective decision making reduces errors by drawing on a wider range of input.

These performance advantages have been shown to stem from the synergy that arises when team members share real-time feedback, pool different skill sets, and engage in collective problem solving (DiBenigno and Kellogg 2014, Page 2019). Such interplay curtails blind spots, encourages scrutiny of multiple viewpoints, and fosters collaborative creativity. By distributing workload and leveraging complementary skills, collaborative teamwork adapts fluidly to shifting requirements, ultimately producing more robust results than isolated contributors could achieve on their own.

Beyond raw performance, a second key rationale for teamwork is the sharing of expertise across functional or disciplinary boundaries (Ayoubi et al. 2017). A central tenet of the knowledge-based view is that specialized knowledge resides in individuals and must be integrated to solve complex problems. Kogut and Zander (1992) show how recombining distinct skill sets can spur innovation, whereas Nickerson and Zenger (2004) emphasize that problem solving often demands multiple domains of expertise working in tandem. Argote (1999), in turn, suggests that teams

are the primary locus of learning and knowledge retention, because members can refine and transfer insights during direct interaction. Additionally, team performance depends not merely on having expertise present, but on the team's capacity to effectively coordinate that expertise through processes of identifying where knowledge exists, recognizing where it is needed, and successfully bringing it to action on tasks (Faraj and Sproull 2000, Orlikowski 2002, Beane and Anthony 2024). In this sense, teamwork serves as on-the-ground conduits of knowledge exchange, bridging cognitive gaps that would otherwise constrain performance.

Recent studies emphasize the importance of distinguishing between functional and industry expertise when understanding collaboration (Kacperczyk and Younkin 2017, Souitaris et al. 2023). Task or functional expertise pertains to the methods and technical principles guiding a given task (Kogut and Zander 1992, Garud 1997), whereas domain expertise focuses on the norms and application contexts that are unique to each sector. Both types of expertise can be relevant for surfacing and implementing innovative solutions effectively (Ayoubi et al. 2026).

The interplay between performance gains and expertise sharing is further magnified by the increasing complexity of modern scientific, technical, and commercial tasks. Wuchty et al. (2007) document a global shift toward greater collaboration across research fields, a trend they link to the expanding breadth of knowledge required to stay at the cutting edge. Jones (2009) frames this as the “burden of knowledge,” showing how deep individual specialization necessitates team-based coordination to integrate fragmented skill sets. In other words, as the volume and sophistication of available knowledge grow, teams have become the indispensable scaffolding to achieve both depth (through specialized experts) and breadth (through interdisciplinary collaboration) in problem solving.

Finally, human collaboration provides critical social and motivational benefits that enhance work satisfaction (Deutsch 1949, Johnson and Johnson 2005, Kozlowski and Bell 2013). Teamwork can create promotive interaction, reducing fear of retaliation and encouraging open participation (Johnson and Johnson 2005). The resulting sense of belonging, collective commitment, and reciprocal support fosters both stronger motivation and greater persistence in challenging tasks.

2.2. Generative AI

Against this backdrop of increasingly team-based knowledge work, GenAI has emerged as a highly promising technology (Dell'Acqua et al. 2023, Noy and Zhang 2023, Peng et al. 2023, Doshi and Hauser 2024, Eloundou et al. 2024, Otis et al. 2024, Boussioux et al. 2025, Brynjolfsson et al. 2025).³ Early studies have focused on GenAI's impact on individual performance,

highlighting gains in productivity, creativity, and decision making. Yet, as the reliance on team-based innovation grows, we need to understand GenAI's influence on collaborative settings—the very context where organizational value is most often created.

GenAI represents a particularly significant development for teamwork because of two distinctive characteristics. Unlike previous waves of technology that primarily automated explicit, codifiable tasks, GenAI can engage with tacit knowledge—the kind of implicit understanding that traditionally could only be shared through direct human interaction (Kogut and Zander 1992, Zander and Kogut 1995, Argote et al. 2021). Additionally, GenAI's ability to engage in natural language dialogue enables it to participate in the kind of open-ended, contextual interactions that characterize effective teamwork, potentially allowing it to serve not just as a tool but as an active participant in collaborative processes (De Freitas et al. 2024, Mollick 2024).

The integration of GenAI into team-based work presents a mix of opportunities and challenges. On one hand, AI can enhance collaborative performance by automating certain tasks and broadening the range of expertise available to team members (Agrawal et al. 2018, Raj and Seamans 2019). It might also enhance collaborative team dynamics and modify the division of labor by expanding the potential performance on certain tasks beyond what humans or AI could achieve on their own (Hoffmann et al. 2024, Choudhary et al. 2025). Finally, AI may also facilitate boundary-spanning across different knowledge domains, drawing a parallel with the effects of earlier technologies (Levina and Vaast 2005, Cattani et al. 2017).

On the other hand, organizational theory cautions that new technologies often require careful integration, lest they destabilize existing routines (March and Simon 1958, Nelson and Winter 1982). Automation may disrupt habitual ways of coordinating tasks (Weber and Camerer 2003). Additionally, when complex tasks are performed by different technologies and individuals, coordination becomes an essential mechanism for managing the gaps created in the division of work (Becker and Murphy 1992, Bailey et al. 2010). A recent laboratory study highlights these potential coordination pitfalls in human–AI partnerships (Dell'Acqua et al. 2025). Even when AI outperforms humans on a specific task, overall team performance declines, reflecting reduced trust and coordination failures. Moreover, technology-driven shifts in roles and expertise may create new silos, limit learning opportunities, or reduce human interaction (Kellogg et al. 2006, Beane 2019, Balasubramanian et al. 2022).

These issues resonate with longstanding concerns that technology can undercut the social aspects of work, thereby lowering human satisfaction, psychological well-being and affecting emotional relationships at

work (Beaudry and Pinsonneault 2010, Dell'Acqua et al. 2025). A growing body of research shows that what finally determines whether new technologies create or destroy value is not only output quality but also the human experience of using them—an angle that is largely absent from productivity-oriented studies. In a thorough ethnographic study, Beane (2019) finds that surgeons confronting robotic systems engage in “shadow learning” to preserve status while coping with heightened anxiety, and these emotions ultimately shape how quickly the technology is mastered. Similarly, field research on algorithmic platforms reveals that frontline employees actively negotiate, resist, or embrace automated controls depending on whether they feel the systems respect their professional autonomy, with workers even engaging in various forms of “algoactivism” (Kellogg et al. 2020). Additionally, shared fears among different organizational groups may hinder information sharing inside organizations, contributing to underperformance in innovation (Vuori and Huy 2016). Together, these pieces of evidence warn that the emotional footprint of a technology can redirect its entire productivity path, making it a core element to understand GenAI's organizational effects.

From this perspective, GenAI represents a further inflection point. Recent meta-analytic evidence suggests that GenAI-based conversational agents can strengthen individuals' social and emotional experience—for example, by providing encouraging, human-like dialogue that reduces distress and fosters well-being, and demonstrating empathetic responses that humans rate as human-like (Li et al. 2023, 2024; Ayers et al. 2023). At the same time, GenAI's characteristics may lead to negative emotional responses. Algorithms that generate responses resembling those of a knowledge worker create opacity that workers struggle to interpret and navigate (Faraj et al. 2018), and “techno-distress” arises when opaque AI recommendations clash with established work norms (Tarafdar et al. 2019). Whether GenAI produces a J-curve dip (Brynjolfsson et al. 2021) or a smooth learning curve therefore hinges in part on the emotions it elicits: positive affect can motivate continued use, whereas confusion or anxiety can stall adoption. Our experiment explores these emotional dynamics alongside performance to ask whether GenAI can become a sustainable teammate that accelerates, rather than derails, long-run organizational learning.

2.3. New Product Development

New product development represents an ideal context for examining how AI affects collaborative work. Unlike abstract creative tasks often studied in laboratory settings, new product development involves real organizational constraints, domain expertise requirements, and measurable commercial outcomes that

mirror the complexity of actual innovation work (Dougherty 1992, Brown and Eisenhardt 1995). This context is particularly valuable because it requires both technical feasibility and market viability which demand the integration of diverse functional expertise that has traditionally justified team-based approaches in organizations (Kogut and Zander 1992, Nickerson and Zenger 2004, Jeppesen and Lakhani 2010, Teodoridis 2018). However, this reliance on teamwork is a double-edged sword, as such cross-functional collaboration faces significant internal barriers in large firms, where different departments often operate with conflicting “thought worlds” that impede successful product innovation (Dougherty 1992).

Moreover, new product development typically involves generating solutions under time pressure with limited resources. This process mirrors key features of an innovation tournament, where organizations must identify the most promising concepts from a broad set of possibilities for further investment and development (Terwiesch and Loch 2004, Terwiesch and Xu 2008, Boudreau et al. 2011). The challenge is compounded by organizational tendencies toward local search and existing routines, which can constrain the exploration of novel solutions precisely when breakthrough thinking is most needed (Nelson and Winter 1982, March 1991), creating a clear opportunity for a cybernetic teammate to potentially broaden the scope of exploration and overcome the challenges posed by these organizational tendencies.

Although recent studies have demonstrated that AI can enhance individual creative performance in controlled settings (Dell'Acqua et al. 2023, Noy and Zhang 2023), to the best of our knowledge, there are no studies examining AI's impact on real-world collaborative team settings. This gap exists even as new theories begin to map out how different “hybrid problem-solving” processes, that is, the specific ways humans and AI collaborate, can lead to systematically different creative outcomes (Raisch and Fomina 2025). Boussioux et al. (2025) provide a critical benchmark, showing that human–AI partnerships can outperform human crowdsourcing in creative problem solving. Our study unpacks the “black box” of this collaborative process in an organizational setting by leveraging the theoretical framework developed by Girotra et al. (2010). We adapt their model to our data, deconstructing performance into a set of distinct levers: the average quality of the ideas generated, the variance in that quality, the effectiveness of selection processes, and the quality of execution in articulating the final chosen idea. This framework is particularly powerful as it incorporates factors like idea variance, shown to be critical for breakthrough innovation (Singh and Fleming 2010), and is motivated by the need for effective selection from a pool of ideas, a concept grounded in

the statistical view of innovation (Dahan and Mendelson 2001). By applying this extended framework to human–AI collaboration in an authentic organizational setting, we can move beyond asking if AI improves team performance to understanding the precise contributions it makes at each stage of the collaborative process.

Innovation research increasingly recognizes that organizations often derive disproportionate value from exceptional outcomes—the very best ideas that may generate outsized returns if implemented. NK-landscape models of organizational search show that on a rugged performance surface, agents who take larger or more varied exploratory steps are more likely to scale the highest peaks, even if their average move is no better than that of more cautious searchers (Levinthal 1997, Rivkin 2000). In innovation contexts, a handful of top ideas can make a significant impact on new product success (Dahan and Mendelson 2001, Girotra et al. 2010, Boudreau et al. 2011). Relatedly, Li et al. (2026) find that algorithms designed for “exploration” rather than mere “exploitation” identify candidates with higher upside potential in hiring contexts, suggesting that AI-augmented collaboration may excel at boundary-spanning exploration that produces rare but disproportionately valuable ideas.

3. Experimental Design

3.1. Empirical Setting

Between May and July 2024, we conducted a field experiment at Procter & Gamble to evaluate how GenAI influences cross-functional new product development.⁴ P&G, a large multinational firm with structured R&D processes and a skilled workforce, provides a useful setting in which to investigate GenAI’s role in innovation-focused knowledge work. With roughly 7,000 R&D professionals worldwide, the firm encompasses end-to-end product development activities, from concept to launch. This breadth of expertise, alongside well-defined organizational routines and substantial operational scope, offers a valuable context through which to examine human collaboration with GenAI in real-world contexts. Over several months, we worked closely with P&G’s leadership to tailor our experimental design, aligning it with the company’s established innovation practices and strategic priorities.

The idea of studying the effects of AI on product innovation tasks at the interplay between commercial and R&D functions originated from several in-depth discussions with the leadership team of the organization. As it often happens in companies of this nature and scale, work at P&G typically occurs in teams and follows structured routines, often involving cross-functional collaboration. This is especially true for innovation activities, for which teams composed of

R&D and commercial representatives are the core units where innovation happens in the company—it is where ideas are generated and the entire innovation funnel begins. Senior executives at P&G emphasized how improving the quality of work at this early stage of the innovation process is crucial for the whole innovation pipeline, producing high-quality “seeds” that can then grow within P&G’s innovation funnel. However, they also reported that coordination frictions—such as finding time to convene representatives of both functions in a meeting, as well as cultural divides between R&D and commercial—could lower the quality of innovation-related activities. The experiment was motivated by the willingness to test how an AI teaming model affects innovation and potentially reduces these frictions.

This setting provides a specific instance where team activity, coordination across functions, and selection processes converge, offering a rich environment to study the impact of AI on collaborative work. By examining how GenAI affects these established collaboration processes, our research provides insights that are directly applicable to the challenges faced by many large organizations in today’s rapidly evolving technological landscape.

3.2. Experimental Approach

This experiment was preregistered prior to data collection.⁵ In line with our theoretical framework, our preregistration focused on two primary dimensions: the effects of introducing AI on performance quality and whether AI could blur the functional boundaries between commercial and R&D professionals. These questions directly correspond to the first two pillars of teamwork—performance and expertise sharing—that frame our investigation.

In presenting our findings, we complement our main analysis with an exploration of the emergent patterns discovered during our investigation. Although performance and expertise dimensions were our primary focus, our examination of emotional responses emerged as a significant factor during the study. This third pillar—sociality, part of the preregistration submission as a variable of interest but with unclear effects—completes our assessment of AI’s potential as a cybernetic teammate. Finally, our exploration of the right tail of performance distribution (breakthrough innovations) emerged as another important dimension, critical to understanding AI’s organizational impact, particularly in innovation contexts.

The experimental design was carefully crafted to mirror P&G’s actual new product development processes, particularly focusing on the early stages where new ideas are generated and initially developed. P&G emphasizes this early seed stage as a crucial element in their entire innovation process. A senior leader at

the company emphasized that “better seeds lead to better trees,” reflecting the importance of high-quality ideation. Through extensive collaboration with P&G over multiple months, we developed a deep understanding of their innovation practices and structured our experiment accordingly. A key insight from this engagement was that early-stage innovation typically involves very small cross-functional teams comprised of commercial and R&D professionals.⁶ We thus mimicked this structure in our experimental design.

The experiment was conducted as a one-day virtual product development workshop, involving 826 participants from P&G’s commercial and R&D functions.⁷ Our analyses focus on 791 of these participants,⁸ who were randomly assigned across four conditions.⁹ Specifically, the four conditions are (1) the control, an individual without AI (Individual No AI); (2) Treatment 1 (T1), a team (R&D + commercial) without AI (Team No AI); (3) Treatment 2 (T2), an individual with AI (Individual + AI); and (4) Treatment 3 (T3), a team (R&D + commercial) with AI (Team + AI). Participants were randomly assigned to these conditions within each of the eight randomization clusters, defined by four business units (Baby Care, Feminine Care, Grooming, and Oral Care) across two geographies (Europe and Americas).¹⁰ Randomization was stratified by business unit and geography to ensure balanced representation across all groups. To preempt social-comparison concerns, we kept participants blind

to the existence of multiple experimental conditions and simply told each person they would experience a personalized version of the workshop. Participants in teams were put in contact only when both attended, ensuring no contamination or selection biases due to participation attrition. Table 1 provides an overview of the participants, indicating a balanced distribution of key functions within P&G. Figure 1 illustrates our 2 × 2 experimental design, with participants randomly assigned to work either individually or in teams, and with or without AI assistance.

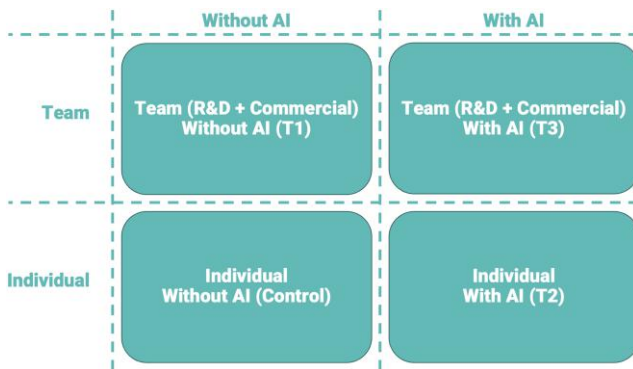
The sample size was determined to ensure sufficient statistical power to detect meaningful differences between conditions, accounting for potential attrition and the nested structure of the data.¹¹ The inclusion of both commercial and R&D functions allows for a comprehensive examination of cross-functional collaboration, a critical aspect of innovation and product development in large consumer goods companies.

The two team conditions (with and without AI) were formed by randomly pairing a commercial and an R&D professional. Collaboration occurred remotely through Microsoft Teams, as is standard practice at P&G, with one team member randomly designated to share their screen and submit the team’s solution.¹² This structure ensured that team members could contribute to and refine their solution in real time, while maintaining a single, coherent workflow for submission. Consequently,

Table 1. Summary Statistics

Panel A: Individual			
	Individual No AI	Individual + AI	Mean diff.
Female	0.578 (0.494)	0.555 (0.497)	−0.023
Male	0.422 (0.494)	0.432 (0.495)	0.010
Band level	2.071 (0.742)	2.065 (0.762)	−0.006
Experience inside company (years)	12.351 (8.293)	11.816 (7.807)	−0.535
R&D specialist	0.604 (0.491)	0.594 (0.493)	−0.010
Use of ChatGPT at work (1–5 Likert)	2.786 (1.126)	2.735 (1.206)	−0.050
Use of ChatGPT personal (1–5 Likert)	2.468 (1.200)	2.529 (1.147)	0.061
Access to ChatGPT at work (yes = 1, no = 0)	0.812 (0.392)	0.800 (0.401)	−0.012
Expectation of AI use at work pre-exp (1–5 Likert)	3.539 (0.951)	3.555 (1.027)	0.016
Individuals	154	155	
Panel B: Team			
	Team No AI	Team + AI	Mean diff.
Female	0.596 (0.492)	0.556 (0.498)	−0.040
Male	0.404 (0.492)	0.444 (0.498)	0.040
Band level	2.000 (0.714)	2.083 (0.734)	0.083
Experience inside company (years)	10.091 (7.616)	10.476 (8.108)	0.385
R&D specialist	0.500 (0.501)	0.500 (0.501)	0.000
Use of ChatGPT at work (1–5 Likert)	2.574 (1.225)	2.615 (1.179)	0.041
Use of ChatGPT personal (1–5 Likert)	2.326 (1.056)	2.480 (1.092)	0.154
Access to ChatGPT at work (yes = 1, no = 0)	0.713 (0.427)	0.746 (0.384)	0.033
Expectation of AI use at work pre-exp (1–5 Likert)	3.430 (1.003)	3.534 (1.021)	0.103
Team participants	230 (115 teams)	252 (126 teams)	

Notes. Standard deviations are in parentheses. Diff., difference; pre-exp, pre-experiment.

Figure 1. Treatment Matrix

Note. This figure displays the 2 × 2 experimental design showing four conditions: individuals and teams working either with or without AI assistance.

our analysis treats each team as a cohesive unit, focusing on overall team performance and AI integration rather than on individual roles within the team structure.

Participants (whether alone or in teams) were assigned tasks within their own business units to develop viable ideas for new products, packaging, communication approaches, or retail execution, among others. All supporting data and processes mirrored what P&G employees would typically use in similar real-world efforts. This design choice enhanced ecological validity by allowing participants to tackle challenges relevant to their day-to-day work. Before random assignment, every participant took part in a brief training session led by a coauthor that reviewed P&G's standard frameworks questions for tackling the task. As all participants were P&G domain experts, this first session served as a refresher and ensured that no group received differential guidance.

The GenAI tool used in the experiment was built on GPT-4 and accessed through Microsoft Azure.¹³ In the AI-enabled conditions (T2 and T3), participants received a one-hour training session on how to prompt and interact with the GenAI tool for Consumer Product Goods-related tasks. One of the authors led this training and provided a PDF with recommended prompts.¹⁴ This standardized approach ensured a uniform baseline of familiarity with the GenAI interface for all AI-enabled participants.¹⁵

In addition to our primary measures of overall performance, expertise sharing, and social interaction, we also collected information on solution novelty, feasibility, and impact as robustness checks. These measures confirm the findings reported in the main text.

3.3. Collected Outcomes

Data collection occurred in multiple stages. Presurvey data were collected to gather individual information about participants. During the product development workshop, all GenAI prompts and responses were recorded, and team interactions were transcribed.

Postsurvey data were also collected, and follow-up interviews were conducted with some participants.

Participant motivation was both intrinsic and extrinsic. First, they enrolled in the study as part of an organizational upskilling initiative to enhance their knowledge about GenAI and its applications in their work. Additionally, a key incentive was the opportunity for visibility: participants were informed that the best proposals would be presented to their respective business unit leaders, offering a chance to showcase their skills and ideas to top management. To maintain fairness and encourage participation across all conditions, rewards for the best proposals were determined within each treatment group (control, individual with AI, etc.). This approach ensured that participants in all conditions had equal opportunities for recognition, regardless of their assigned experimental group.

4. Empirical Strategy

4.1. Analytical Approach

Our empirical analysis primarily relies on regression analysis to estimate the causal effect of AI adoption and team configuration on various outcome measures. Our main specification takes the following form for a given solution generated i :

$$Y_i = \beta_0 + \beta_1 \text{TeamNoAI}_i + \beta_2 \text{AloneAI}_i + \beta_3 \text{TeamAI}_i + \gamma \text{Controls}_i + \delta \text{FE}_i + \epsilon_i,$$

where Y_i represents different outcome variables that we examine in our analysis. Each outcome captures a distinct dimension of performance, expertise and collaboration that we investigate to understand the multifaceted impact of AI adoption and team configuration on work processes and outputs. The baseline category is individuals without AI. We describe these outcome variables in detail in Section 4.2 below.

The term Controls_i includes a list of preexperimental features including demographic and professional characteristics, and FE_i includes day and business unit fixed effects.

We estimate three variants of this model. Model 1 includes only the treatment indicators. Model 2 includes only fixed effects for business unit and date of participation. Model 3 adds controls including band level, years of experience in the company, gender, and prior AI usage both at work and for personal purposes.¹⁶ Throughout our analysis, we use robust standard errors to account for potential heteroskedasticity.¹⁷

Beyond these direct comparisons to the baseline, we conduct additional analyses comparing outcomes across treatment conditions. Of particular interest are the comparisons between the two team conditions (team without AI versus team with AI) and between the two AI-enabled conditions (alone with AI versus team with AI). These additional comparisons help us

understand both the value of AI in team settings and the complementarity between AI and teamwork. Whenever relevant, we report the p -values for these comparisons at the bottom of our regression tables and discuss their implications in the text.

4.2. Dependent Variables

Our primary outcome measure is *Quality*, which captures the overall quality of submitted solutions on a scale from 1 to 10. These quality scores were assigned by human expert evaluators with backgrounds in both business and technology, who independently assessed each solution. The evaluators were blind to the conditions of the experiment and the profile of the submitters. We standardized these scores based on the control group (individuals working alone without AI), resulting in scores that represent standard deviations from the control group mean. During the same evaluation process, experts also assessed two additional key dimensions of the solutions: *Novelty* and *Feasibility*.¹⁸ *Novelty* measures the degree of innovation and originality in the submitted solutions on a scale from 1 to 10, whereas *Feasibility* evaluates how practical and implementable the solutions are, also on a 1–10 scale. These dimensions were evaluated simultaneously with the overall quality assessment, providing a comprehensive evaluation of each solution's merits.¹⁹

These innovation outcomes are grounded in the literature (e.g., Boudreau et al. 2016, Lane 2023) and are also used extensively by P&G. On average, each solution received more than three independent evaluations, though the exact number varies across solutions. This multiple-evaluation approach helps ensure the robustness of our quality measurements.

We also analyze other performance measures such as *Solution Length* and *Expected Quality*. *Solution Length* measures the total number of words in the solutions submitted by participants. This variable helps us understand how AI and team configuration affect the comprehensiveness and detail level of submitted solutions.

Expected Quality is a binary variable based on survey responses, where participants indicated whether they expected their solution to rank in the top 10% (one) or not (zero). Participants answered this question after submitting their final solution. This measure helps us understand how different working configurations affect participants' confidence and self-assessment of their performance.

In addition to performance metrics, we capture how expertise is configured and deployed. Specifically, we categorize participants based on their domain of knowledge (*R&D* or *Commercial*) and their functional experience embodied in whether product development is a *Core Job* responsibility (i.e., employees who regularly engage in new product initiatives) or a

non-core-job role (i.e., individuals in the same business unit but involved less frequently in new product innovation). This dichotomy provides insight into how prior knowledge and domain familiarity might interact with AI or team structures. Additionally, we measured the degree of *Technicality* of a solution using a one-to-seven Likert score assigned by the same human evaluators assessing solution quality, where higher values indicate more technically oriented ideas. Conversely, lower values suggest commercially oriented, market-focused concepts.

Finally, we measure changes in participants' self-reported emotional states before and after completing the task through two composite measures. Positive emotions combine participants' reported levels of *Enthusiasm*, *Energy*, and *Excitement*, whereas negative emotions aggregate feelings of *Anxiety*, *Frustration*, and *Distress*.²⁰ Each component is measured on a scale from one to seven, and standardized based on the control group mean and standard deviation. Both measures are calculated as the difference between post-task and pretask responses: that is, we measure emotional change from baseline levels established at the beginning of the task, not absolute emotional states. Because baseline measurements were taken after treatment assignment but before task engagement, any initial emotional reactions to assignment conditions would already be incorporated in these baselines, making the observed positive emotional shifts attributable to the actual experience of working with AI.²¹

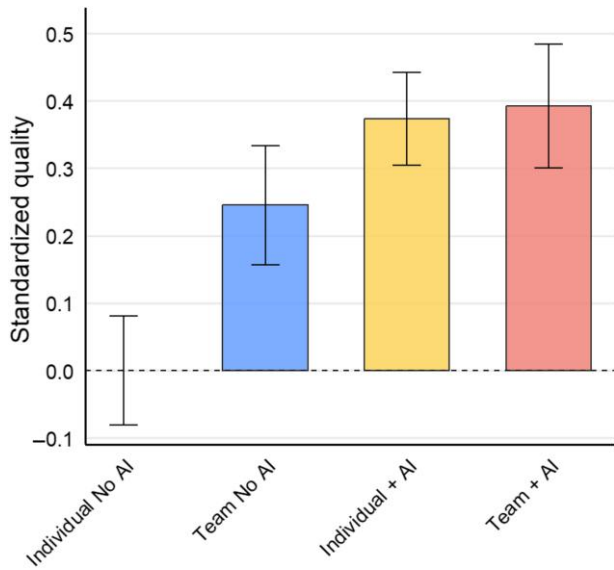
5. Results

5.1. Performance

Figure 2 provides instructive insights into the quality of solutions across different groups. It displays average quality scores, showing the relative performance of AI-treated versus non-AI treated groups is significantly higher. The distributions of these quality scores, shown in Figure 3, reveal that while both teams without AI and individuals with AI significantly outperform the control group, their quality distributions are remarkably similar, providing further evidence that AI can replicate key performance benefits of teamwork. Table 2 quantifies these quality differences through regression analysis. Teams without AI show a quality improvement of 0.24 standard deviations over individuals without AI ($p < 0.05$), highlighting the traditional benefits of collaboration. This replication of traditional team benefits serves as an important validation of our experimental setting, confirming that teams function as expected in real organizational contexts, as well as confirming P&G's new product development experience.

The impact of AI is greater: individuals with AI demonstrate a 0.37 standard deviation increase ($p < 0.01$), and teams with AI show a 0.39 standard deviation

Figure 2. Average Solution Quality

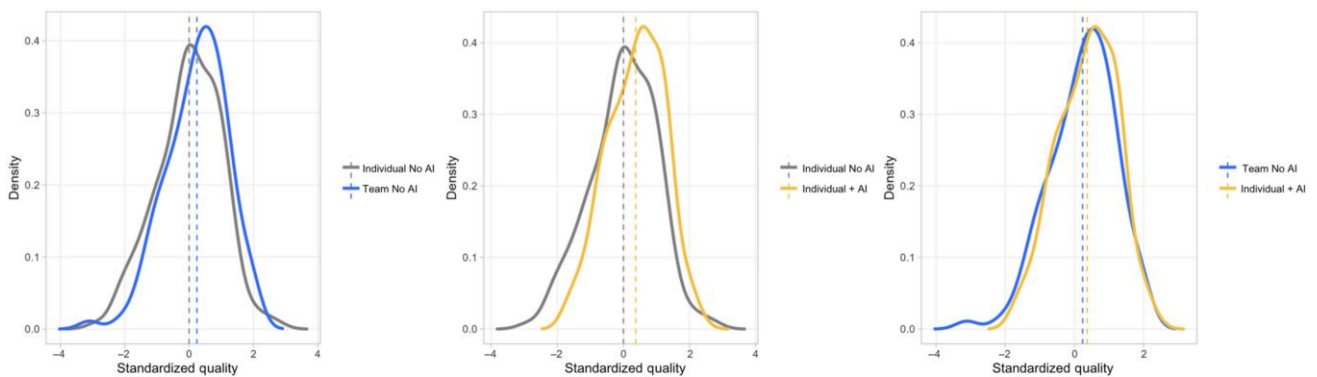


Note. This figure displays the average quality scores for solutions across different groups, showing the relative performance of AI-treated versus non-AI-treated groups with standard errors.

improvement ($p < 0.01$). These effects remain robust across all specifications. The data reveal a hierarchy in solution quality across different working configurations. Individuals working alone without AI assistance produced the lowest-quality solutions on average. Teams working without AI showed a modest improvement over individuals. The introduction of AI led to notable performance changes: individuals working with AI performed at a level comparable to teams without AI, suggesting that AI-enabled individuals can match the output quality of traditional human teams, effectively substituting for team collaboration in certain contexts.

Finally, as has been the case with individual workers, we see AI leading to more developed outcomes.

Figure 3. Pairwise Density Comparisons



Notes. These figures illustrate the pairwise comparisons of solution quality distributions across different experimental conditions. The left panel compares solutions between individuals and teams working without AI assistance. The middle panel shows the quality distribution between individuals working alone with and without AI assistance. The right panel compares solutions between teams without AI and individuals with AI assistance.

Table 2. Solution Quality (Standardized)

	Quality	Quality	Quality
<i>Team No AI</i>	0.245** (0.120)	0.262** (0.122)	0.307** (0.131)
<i>Individual + AI</i>	0.373*** (0.106)	0.386*** (0.108)	0.370*** (0.107)
<i>Team + AI</i>	0.392*** (0.122)	0.404*** (0.123)	0.463*** (0.139)
<i>Team + AI = Team No AI</i>	$p = 0.242$	$p = 0.254$	$p = 0.216$
Fixed effects		X	X
Controls			X
Control mean	0.000 (0.081)	-0.173 (0.173)	0.306 (0.228)
Observations	550	550	550
Adjusted R^2	0.023	0.023	0.048

Notes. The p -values for the t -tests comparing *Team + AI* and *Team No AI* are reported. Fixed effects and controls are as discussed in the text. ** $p < 0.05$; *** $p < 0.01$.

Whereas teams without AI produced solutions only marginally longer than individual controls, the introduction of AI led to significantly longer outputs. As shown in Table 3, these large effects persist across all specifications.

5.2. Expertise

We now turn to how AI impacts how team expertise is leveraged in the new product development task. We start by examining the heterogeneity of the results across workers who have different familiarity with this type of task, as shown in Figure 4 and the corresponding Table 4. Figure 4 splits our sample between employees for whom product development is a core job task (left panel; core job) and employees who are less familiar with new product development (right panel; non-core job), comparing their performance across our experimental conditions.²²

Table 3. Solution Length

	Length	Length	Length
<i>Team No AI</i>	30.456 (27.419)	56.746* (30.865)	57.184+ (38.673)
<i>Individual + AI</i>	504.507*** (42.963)	511.568*** (45.206)	503.833*** (45.081)
<i>Team + AI</i>	543.745*** (42.328)	556.997*** (43.737)	551.578*** (51.989)
Fixed effects		X	X
Controls			X
Control mean	381.422	306.565	336.197
Observations	550	550	550
Adjusted R ²	0.317	0.337	0.344

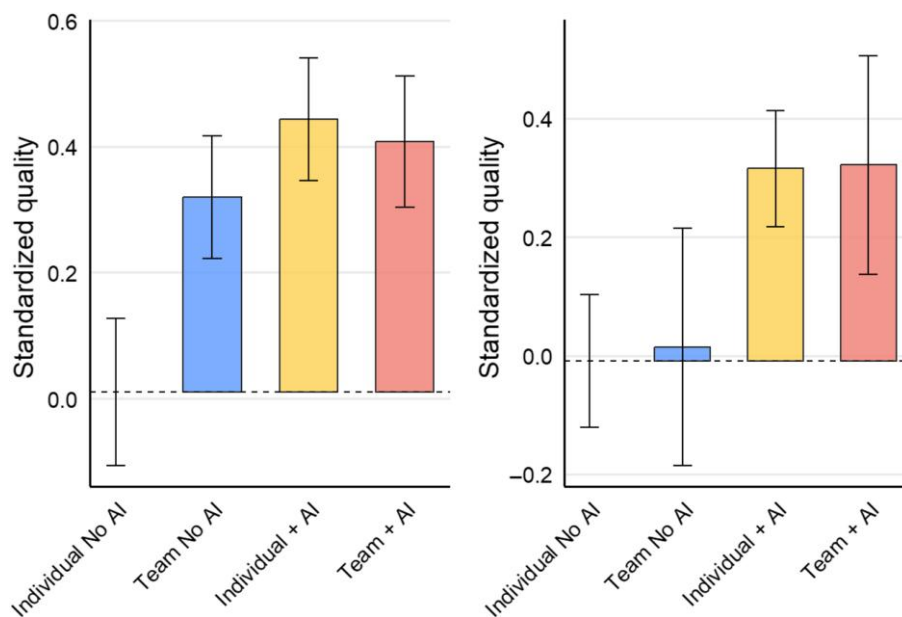
Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text.
 + $p < 0.2$; * $p < 0.1$; *** $p < 0.01$.

The results are particularly noteworthy for non-core-job employees. Without AI, non-core-job employees working alone performed relatively poorly. Even when working in teams, non-core-job employees without AI showed only modest improvements in performance. However, when given access to AI, non-core-job employees working alone achieved performance levels comparable to teams with at least one core-job employee. This suggests that AI can effectively substitute for the expertise and guidance typically provided by team members that are familiar with the task at hand. This pattern demonstrates AI's potential to improve access to expertise within organizations, extending prior work on individual knowledge

workers (e.g., Dell'Acqua et al. 2023, Brynjolfsson et al. 2025). AI allows less experienced employees to achieve performance levels that previously required either direct collaboration or supervision by colleagues with more task-related experience.

Figure 5 illustrates the difference in idea generation between commercial and technical participants, with and without AI assistance. The left graph shows participants working alone without AI. In this scenario, commercial participants (green) demonstrate a higher likelihood of proposing less technical ideas, as indicated by their distribution toward lower values on the x -axis. In contrast, technical participants (yellow) tend to suggest more technically oriented ideas, clustering toward higher x -axis values. The right graph depicts participants working with AI assistance. Notably, the distinction between commercial and technical participants disappears in this scenario. The distribution of both groups appears similar across the x -axis, suggesting that AI assistance leads these groups to propose ideas of a similar level of technicality. Figure 5 illustrates a shift in idea generation patterns with the introduction of AI. Without AI assistance, participants tended to generate ideas closely aligned with their professional backgrounds. However, when aided by AI, this distinction largely disappeared. Both commercial and technical participants generated a more balanced mix of ideas, spanning the commercial/technical spectrum. Moreover, quality scores did not significantly vary based on a solution's technical orientation, indicating that these effects did not come at the cost of solution

Figure 4. Average Solution Quality: Core Jobs vs. Not



Note. This figure displays the average quality scores for solutions across different groups, separating between participants who are more familiar with this type of task (on the left) and participants less familiar with it (on the right) with standard errors.

Table 4. Solution Quality by Familiarity with the Type of Task (Standardized)

	Quality					
	Noncore jobs			Core jobs		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
<i>Team No AI</i>	0.023 (0.228)	0.026 (0.240)	−0.132 (0.248)	0.309** (0.152)	0.328** (0.151)	0.377** (0.165)
<i>Individual + AI</i>	0.324** (0.149)	0.356** (0.151)	0.360** (0.156)	0.433*** (0.152)	0.457*** (0.150)	0.457*** (0.153)
<i>Team + AI</i>	0.330+ (0.213)	0.299+ (0.212)	0.203 (0.253)	0.397** (0.157)	0.386** (0.157)	0.455** (0.179)
Fixed effects		X	X		X	X
Controls			X			X
Control mean	−0.009 (0.112)	−0.194 (0.258)	0.382 (0.336)	0.010 (0.117)	−0.143 (0.232)	0.311 (0.317)
Observations	218	218	218	332	332	332
Adjusted R^2	0.014	0.009	0.032	0.019	0.040	0.062

Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text.

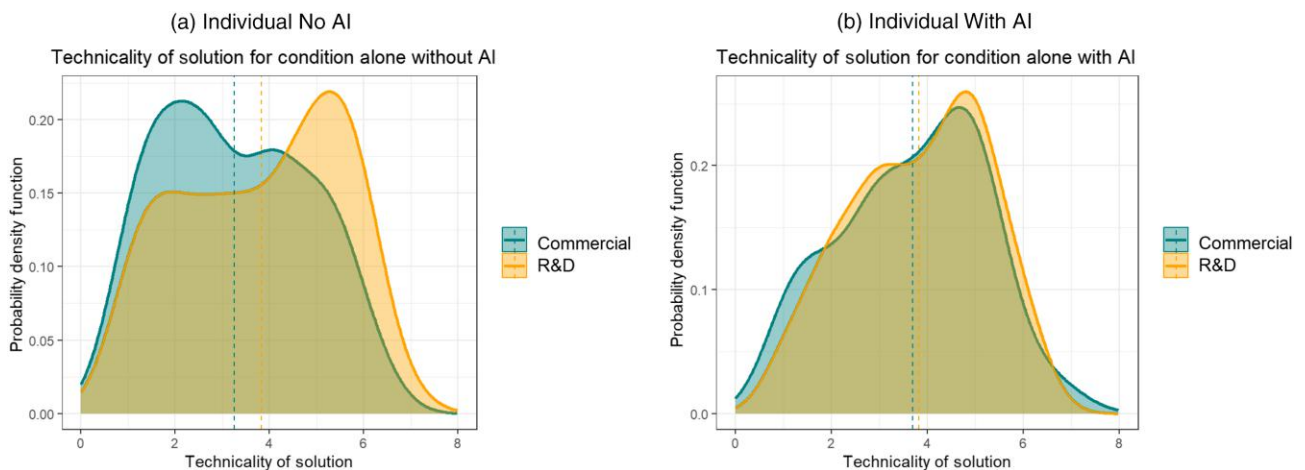
+ $p < 0.2$; ** $p < 0.05$; *** $p < 0.01$.

effectiveness. By leveraging AI, participants effectively expanded their problem-solving horizons, demonstrating AI's potential to foster more holistic and interdisciplinary thinking.

5.3. Sociality

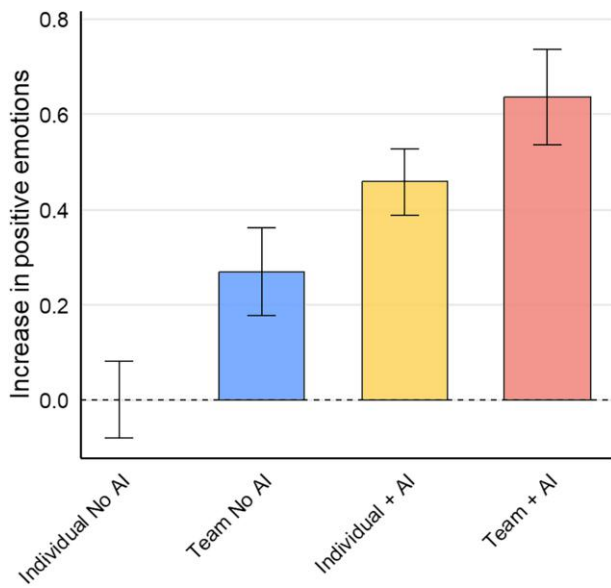
Finally, we find that AI integration leads to enhanced positive emotional experiences. Figures 6 and 7 present emotional responses across groups, illustrating that participants using AI reported significantly higher levels of positive emotions (excitement, energy, and enthusiasm) and lower levels of negative emotions (anxiety and frustration). Tables 5 and 6 confirm these results. Specifically, individuals with AI showed a

0.457 standard deviation increase in positive emotions ($p < 0.01$) compared with the control group, and teams with AI demonstrated an even larger 0.635 standard deviation increase ($p < 0.01$). Simultaneously, both individuals and teams using AI reported significant decreases in negative emotions (−0.233 and −0.235 standard deviations respectively, $p < 0.05$). This pattern of emotional responses provides further evidence of AI's effectiveness as a teammate. Without AI assistance, individuals working alone show lower positive emotional responses compared with those working in teams, reflecting the traditional psychological benefits of human collaboration. However, individuals using AI report positive emotional responses that match or

Figure 5. Degree of Solution Technicality for Individuals

Notes. These figures illustrate the difference in idea generation between commercial and technical participants, with and without AI assistance. In both graphs, blue represents commercial participants and yellow represents technical participants. The x -axis indicates the commercial nature of ideas, with higher values representing more technically oriented suggestions.

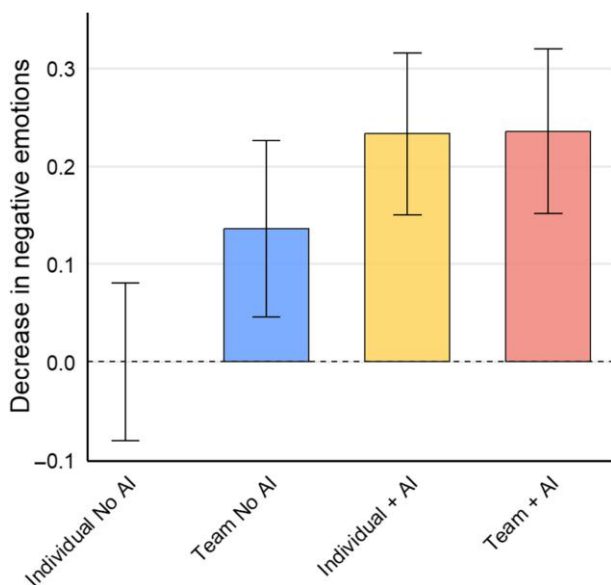
Figure 6. Evolution of Positive Emotions During the Task



Notes. This figure presents the difference in self-reported positive emotions among participants before and after the task, comparing AI-treated and non-AI-treated groups to examine the emotional impact of AI on teamwork with standard errors. Positive emotions are answers to questions about enthusiasm, energy, and excitement. Higher numbers indicate stronger emotional responses.

exceed those of team members working without AI. This suggests that AI can substitute for some of the emotional benefits typically associated with teamwork, serving as an effective collaborative partner even in

Figure 7. Evolution of Negative Emotions During the Task



Notes. This figure presents the reduction in self-reported negative emotions among participants before and after the task, comparing AI-treated and non-AI-treated groups to examine the emotional impact of AI on teamwork with standard errors. Negative emotions are answers to questions about anxiety, frustration, and distress. Higher numbers indicate negative emotions decreased.

Table 5. Evolution of Self-Reported Positive Emotions Before and After the Task (Standardized)

	Positive emotions	Positive emotions	Positive emotions
<i>Team No AI</i>	0.269** (0.124)	0.254** (0.126)	0.257* (0.137)
<i>Individual + AI</i>	0.457*** (0.107)	0.475*** (0.106)	0.485*** (0.106)
<i>Team + AI</i>	0.635*** (0.131)	0.635*** (0.129)	0.666*** (0.153)
Fixed effects		X	X
Controls			X
Control mean	0.000	-0.315	0.012
Observations	533	533	533
Adjusted R ²	0.050	0.064	0.070

Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

individual work settings. At the same time, it is important to recognize that not all negative emotions in teams are necessarily detrimental. A certain degree of creative tension and disagreement can stimulate deeper exploration and higher-quality ideas over time, suggesting that the reduction of interpersonal friction we observe here may not always translate into long-term creative gains and may in fact be harmful (Jonassen et al. 2026).

These emotional responses correlate with participants' evolving expectations about AI use. As shown in Tables 7 and 8, participants who reported larger increases in their expected future use of AI also reported more positive and fewer negative emotions during the task. Although this correlation cannot definitively establish causality, it suggests an interesting relationship between positive experiences with

Table 6. Evolution of Self-Reported Negative Emotions Before and After the Task (Standardized)

	Negative emotions	Negative emotions	Negative emotions
<i>Team No AI</i>	-0.136 (0.124)	-0.094 (0.121)	-0.006 (0.141)
<i>Individual + AI</i>	-0.233** (0.117)	-0.247** (0.116)	-0.263** (0.117)
<i>Team + AI</i>	-0.235** (0.118)	-0.221* (0.116)	-0.157 (0.138)
Fixed effects		X	X
Controls			X
Control mean	0.000 (0.082)	0.166 (0.166)	0.068 (0.252)
Observations	530	530	530
Adjusted R ²	0.005	0.022	0.031

Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text.

* $p < 0.1$; ** $p < 0.05$.

Table 7. Average Evolution of Self-Reported Positive Emotions Before and After the Task Based on Expectation of Use of AI at Work

	Without AI (control)			With AI (treatment)		
	Positive emotions	Positive emotions	Positive emotions	Positive emotions	Positive emotions	Positive emotions
<i>Diff. in expected use of GenAI</i>	0.297* (0.171)	0.231+ (0.178)	0.140 (0.182)	0.678*** (0.248)	0.701*** (0.234)	0.638*** (0.243)
Fixed effects		X	X		X	X
Controls			X			X
Control mean	−0.992	−1.606	−1.083	0.013	−0.931	0.992
Observations	262	262	262	271	271	271
Adjusted R^2	0.007	0.025	0.036	0.029	0.059	0.086

Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text. Diff., Difference.

+ $p < 0.2$; * $p < 0.1$; *** $p < 0.01$.

AI and anticipated future engagement with the technology.

5.4. Beyond Aggregate Performance: Process and Extremes

5.4.1. Decomposing AI's Contributions. To understand the specific mechanisms through which AI affects collaborative performance, we leverage our experimental design's multistage structure, which allows us to decompose the innovation process into its core statistical levers. Participants first generated five initial ideas, then selected the one they wanted to proceed with, and finally developed that chosen idea into a detailed solution. By constraining idea quantity to five across all conditions—an explicit design choice that suppresses an additional channel through which AI could provide an advantage—we can isolate AI's impact on the core levers identified in our theoretical framework: average idea quality, quality variance, and selection effectiveness.

Figure 8 reveals where AI exerts its primary collaborative influence across the innovation process. Panel (a) demonstrates that AI enhances the average quality

of generated ideas, with both AI-enabled conditions showing marked improvements over their non-AI counterparts. Panel (b) examines selection effectiveness. We measure it as the probability of choosing the highest-quality idea to go forward with. This selection process reveals an interesting pattern: teams without AI demonstrate the strongest capability at identifying their best idea from their portfolio of five, correctly selecting their highest-quality concept approximately 50% of the time compared with roughly 37% for AI-enabled conditions. Interestingly, panel (c) shows that AI's improvement in average quality occurs across the full distribution of idea quality—the range between highest and lowest-quality ideas remains extremely similar across all conditions, indicating that AI elevates the entire quality spectrum rather than constraining variance or eliminating creative extremes. However, panel (d) shows that despite the selection disadvantage observed in panel (b), AI conditions still produce higher-quality selected ideas in absolute terms. This occurs because AI's boost to average idea quality (panel (a)) more than compensates for any modest reduction in selection accuracy, resulting in

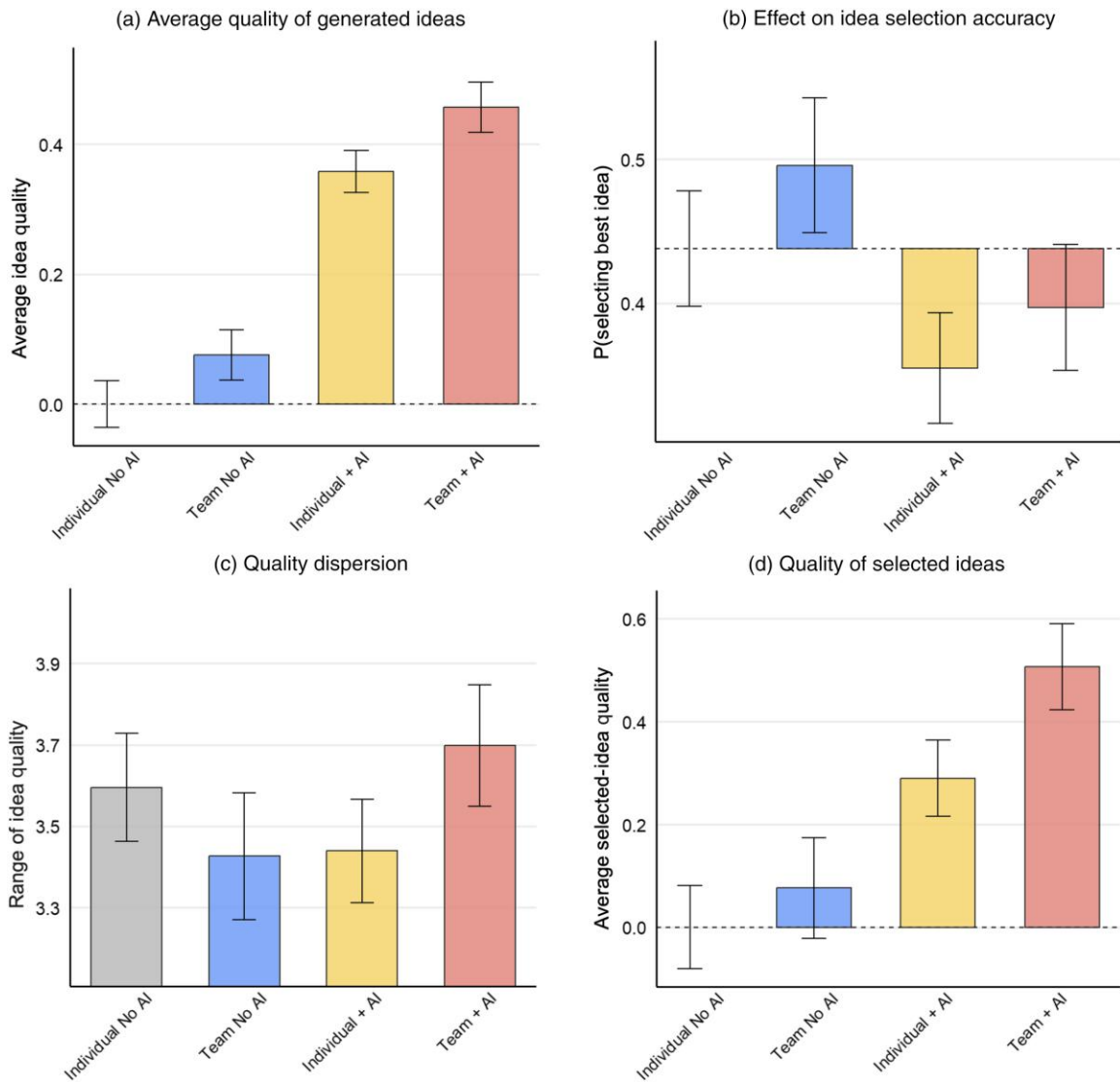
Table 8. Average Evolution of Self-Reported Negative Emotions Before and After the Task Based on Expectation of Use of AI at Work

	Without AI (control)			With AI (treatment)		
	Negative emotions	Negative emotions	Negative emotions	Negative emotions	Negative emotions	Negative emotions
<i>Diff. in expected use of GenAI</i>	−0.270* (0.137)	−0.240* (0.144)	−0.170 (0.154)	−0.581*** (0.190)	−0.607*** (0.188)	−0.663*** (0.201)
Fixed effects		X	X		X	X
Controls			X			X
Control mean	−0.134	0.122	0.109	−0.449**	0.110	0.880
Observations	259	259	259	271	271	271
Adjusted R^2	0.007	0.032	0.071	0.023	0.028	0.077

Notes. Standard errors are in parentheses. Fixed effects and controls are as discussed in the text. Diff., Difference.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure 8. Decomposition of AI's Impact on the New Product Development Process



Notes. Panel (a) shows the average quality of ideas generated across experimental conditions, standardized relative to the control group (Individual No AI). Panel (b) displays the probability that participants selected their highest-quality idea to develop further. Panel (c) displays the average gap between highest- and lowest-quality ideas within each portfolio. Panel (d) shows the average quality of selected ideas; although AI conditions show higher absolute quality, this reflects their elevated baseline rather than improved selection capability. All error bars represent standard errors.

superior final outcomes even when participants are slightly less effective at identifying their strongest concepts.

Panel (b) shows that participants working without AI appear to have been better at identifying their own best ideas than those who worked with AI. This pattern suggests that AI assistance may subtly undermine the evaluative judgment needed to select the most promising concepts from a generated set. Several mechanisms could be at play. For one, the sycophantic tendencies sometimes observed in LLMs may have reinforced participants' confidence in their initial ideas (Randazzo et al. 2025a). The validating nature of AI feedback may itself erode critical engagement: unlike

human teammates, who naturally introduce friction and dissent, AI tends to affirm. That affirmation may feel productive in the moment while diminishing participants' evaluative judgment—a dynamic consistent with evidence that AI overreliance can cause users to exert less effort (Dell'Acqua 2022). Additionally, when ideas are developed with AI assistance, they may be less deeply internalized by their human collaborators, making critical evaluation more challenging—and recent work suggests that AI-generated explanations can suppress independent human judgment in evaluations (Lane et al. 2026). That being said, AI's boost to idea quality more than compensates for this selection effect, yielding superior final outcomes overall.

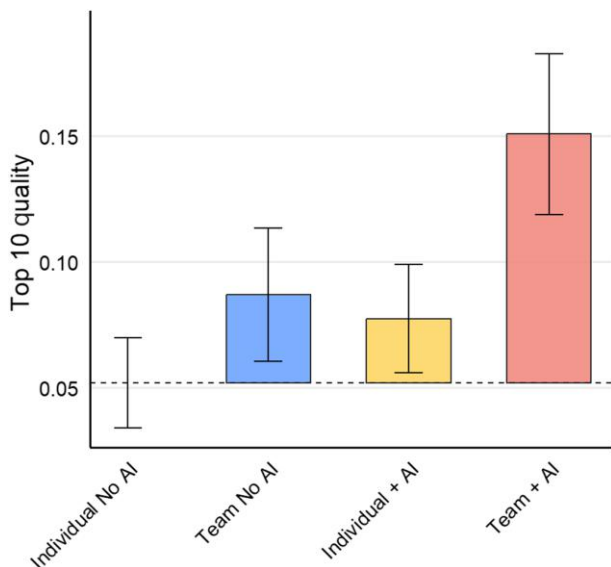
Overall, these results illuminate AI's primary mechanism as a collaborative partner: it functions as a quality amplifier rather than a decision enhancer. AI consistently elevates the baseline quality of creative output while also preserving the natural variance that drives breakthrough innovation. Interestingly, human teams without AI demonstrate a small advantage in selection accuracy, suggesting that human-to-human collaboration may offer unique benefits for evaluative judgment. This confirms GenAI's unclear potential as a decision maker for innovation selection (Csaszar et al. 2024, Doshi et al. 2025, Lane et al. 2026), noting that AI-aided participants may have not used AI for the selection of ideas.

5.4.2. Exceptional Performance Measures. Although the decomposition analysis reveals how AI affects average performance across different stages of the innovation process, many organizations place disproportionate emphasis on exceptional outcomes that can reinforce their competitive position. To explore whether AI facilitates standout solutions, we examine the likelihood of generating top-tier innovations across our experimental conditions.

We developed additional metrics capturing top-tier performance. We created a binary measure called *Top 10% Solutions*, which equals one if a solution's quality score (on a 1–10 scale) ranked in the highest decile across all submissions in the sample, and zero otherwise. By isolating these top performers, we can assess the extent to which AI-enabled conditions and team configurations produce exceptionally high-quality innovations.

Figure 9 highlights the extent to which AI improves innovative performance. Both individuals and teams

Figure 9. Top 10% Solutions



Note. This figure displays the proportion of top 10% solutions across different treatments with standard errors.

Table 9. Probability of Being Rated Top 10% of Quality Scores

	Top quality	Top quality	Top quality
<i>Team No AI</i>	0.037 (0.033)	0.045 ⁺ (0.034)	0.054 ⁺ (0.041)
<i>Individual + AI</i>	0.019 (0.029)	0.029 (0.029)	0.030 (0.029)
<i>Team + AI</i>	0.092** (0.037)	0.098** (0.038)	0.112** (0.045)
<i>Team + AI = Team No AI</i>	$p = 0.190$	$p = 0.207$	$p = 0.175$
<i>Team + AI = Individual + AI</i>	$p = 0.061$	$p = 0.077$	$p = 0.069$
Fixed effects		X	X
Controls			X
Control mean	0.058	−0.040	0.025
Observations	550	550	550
Adjusted R^2	0.008	0.010	0.003

Notes. The p -values for the t -tests comparing *Team + AI* with *Team No AI* and *Individual + AI* are reported. Fixed effects and controls are as discussed in the text.

⁺ $p < 0.2$; ^{**} $p < 0.05$.

using AI were more likely to generate solutions ranking in the top 10% of all submissions. Specifically, as quantified in Table 9, teams with AI were 9.2 percentage points more likely to produce solutions in the top decile compared with the control mean of 5.8%, which corresponds to roughly three times more chances of being in the top decile of solutions. Although individuals with AI show a small positive effect, this effect is not statistically significant, suggesting that the combination of AI and teamwork might be particularly powerful for achieving exceptional performance.

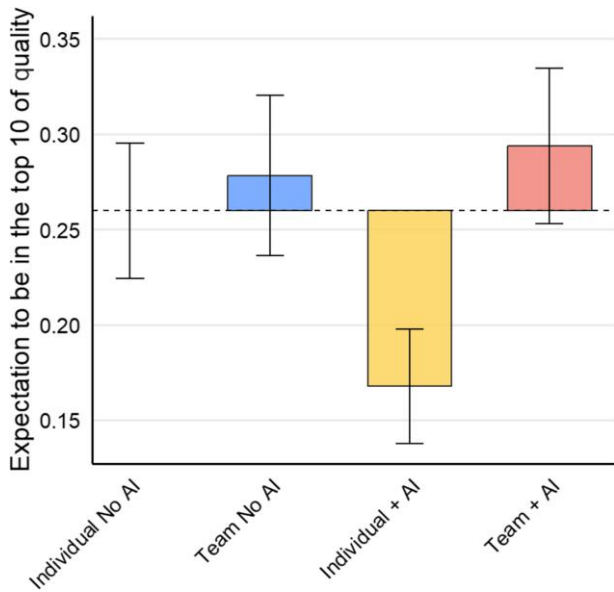
These patterns indicate that AI, particularly when combined with teamwork, does not just improve average performance but also increases the likelihood of producing the kind of breakthrough solutions that drive organizational success.

5.5. Additional Analyses

5.5.1. Expected Quality. We captured *Expected Quality*—a self-reported binary variable indicating whether participants believed their solution would be in the top 10% or not. Participants answered this question immediately after submitting their final solution. Interestingly, although objective performance improved, participants using AI were actually less confident about their solutions. As shown in Figure 10, AI-enabled participants were 9.2 percentage points less likely to expect their solutions to rank in the top 10% compared with the control group ($p < 0.05$), suggesting a disconnect between actual and perceived performance.

5.5.2. Human Team Collaboration. Figure 11 shows the distribution of solution types, ranging from technically focused to market-focused approaches. Without

Figure 10. Perceived Likelihood of Top 10 Percent Placement by Treatment Group



Notes. This table shows the percentage of participants in each treatment group who expected their solution to rank among the top 10 percent. It reflects participants' confidence in their solutions across different conditions with standard errors.

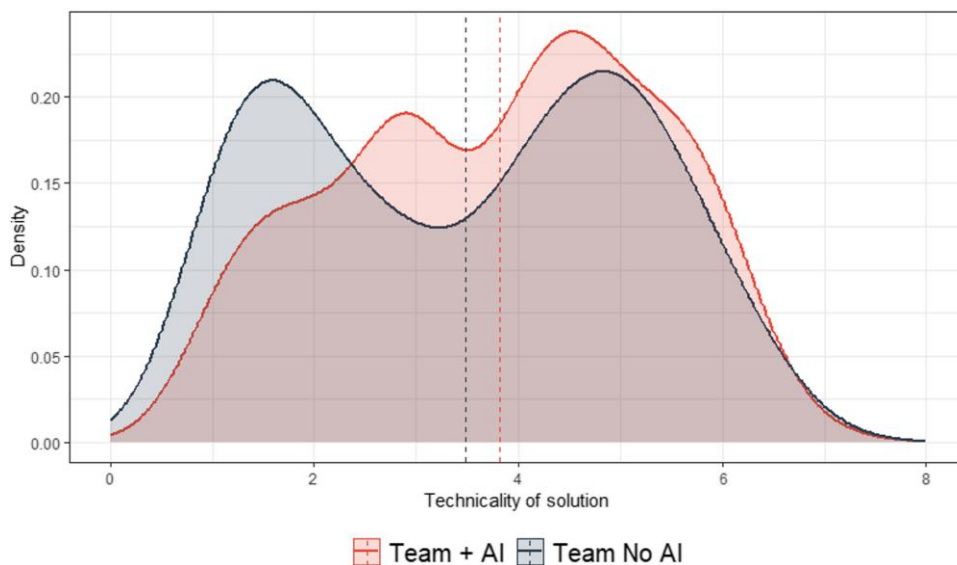
AI, teams exhibit a clear bimodal distribution (bimodality coefficient = 0.564), suggesting that solutions tend to cluster around either technical or commercial orientations, likely reflecting the dominant perspective of the more influential team member. In contrast, AI-enabled teams show a more uniform, unimodal distribution (bimodality coefficient = 0.482), while maintaining similar

overall levels of technical content. This moderate shift from bimodality to unimodality, while preserving the range of technical depth, suggests that AI helps reduce dominance effects in team collaboration. Overall, AI appears to facilitate more balanced contributions from both technical and commercial perspectives.

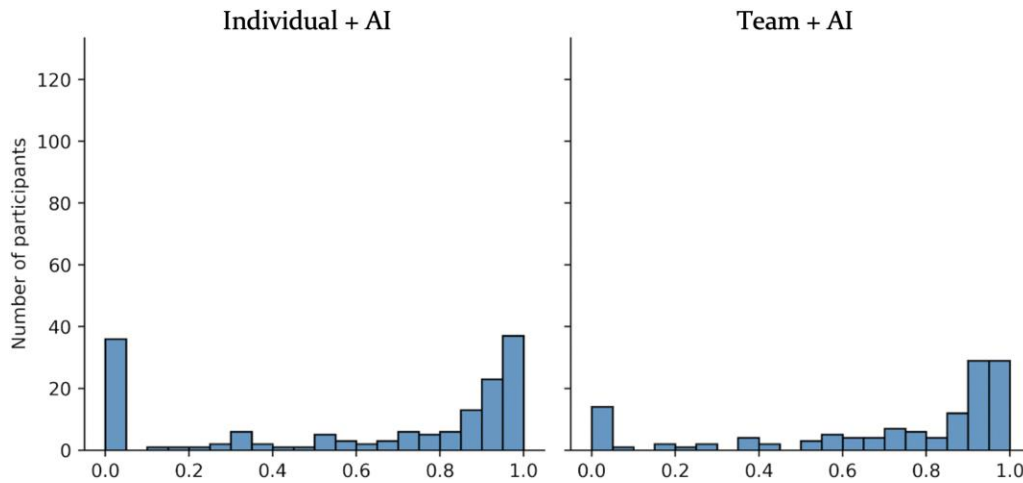
5.5.3. Patterns of AI Use. Our data also allowed us to assess the extent to which teams actually used the AI in their work. To assess the extent of AI utilization in solution generation, we analyzed the retention rate of AI-generated content in participants' final submissions. Our retention measure quantifies the percentage of sentences in the submitted solutions that were originally produced by AI, with a threshold of at least 90% similarity. This metric excludes sentences that were part of the initial human-authored prompts, focusing solely on AI-generated content. Figure 12 illustrates the distribution of retention rates for both individual and group AI conditions.

The retention analysis reveals an interesting pattern relating to AI reliance among participants. For both individuals and groups using AI, we observe a significant skew toward high retention rates, with a substantial proportion of participants retaining more than 75% of AI-generated content in their final solutions. This suggests that many participants heavily leveraged AI capabilities in crafting their responses. However, high retention rates do not necessarily indicate passive AI adoption—participants may engage extensively with the tool through iterative prompting, validation of responses, critical evaluation, and incorporation of domain expertise in their prompting strategy.

Figure 11. Degree of Solution Technicality for Teams



Notes. These figures illustrate the difference in idea generation for teams. Dark blue represents Team No AI, and red represents Team + AI. The x-axis indicates the commercial nature of ideas, with higher values representing more technically oriented suggestions.

Figure 12. Retention of AI-Aided Solutions

Notes. This figure shows the distribution of AI-generated content retained in final solutions for AI-treated participants (individuals and teams). The retention rate represents the proportion of sentences in submitted solutions that were originally produced by AI (with at least 90% similarity), excluding content from initial human prompts.

Interestingly, the distribution also shows a nontrivial percentage of participants with zero retention. These cases represent participants who engaged with AI for ideation, brainstorming, or validation purposes rather than direct solution generation. This polarized distribution points to two distinct patterns of AI usage: one where participants heavily rely on AI-generated content for their final solutions, and another where AI serves primarily as a collaborative tool for ideation and refinement rather than direct content generation, confirming the broad variety in the style of use of GenAI by workers (Randazzo et al. 2025b).

6. Discussion

Our study offers insights about the potential impact of GenAI on team collaboration in the workplace, with implications for both theory and practice. Our findings suggest that AI integration does more than augment existing work processes and may also affect the nature of collaboration and expertise in organizational settings. Our results begin by confirming traditional assumptions about team effectiveness—teams without AI demonstrated modestly better performance (0.24 standard deviation, which represents around 6.3% improvement on the final outcome) compared with individuals working alone, reflecting the traditional benefits of cross-functional collaboration. However, the introduction of AI has the potential to gradually modify this performance landscape. Individuals working with AI showed a 0.37 standard deviation performance increase (equating to around 9.6% improvement) over the baseline of working alone without AI.²³ This finding suggests that AI can effectively substitute for certain collaborative functions, acting as a genuine teammate by granting individuals access to the varied expertise

and perspectives traditionally provided by team members. Teams augmented with AI showed similar levels of improvement (0.39 standard deviations, or around 10.2% improvement, over baseline): their performance was not significantly different from that of individuals using AI. This pattern suggests that AI's immediate impact appears to stem more from its capacity to bolster individual cognitive capabilities than affecting human-to-human collaboration.

Building on these performance patterns, perhaps our most noteworthy finding concerns AI's role in blurring professional expertise boundaries. Organizational theory has long emphasized the importance of specialized knowledge and clear functional boundaries. Our results suggest AI is starting to disrupt this paradigm. Without AI, we observed clear professional silos—commercial specialists submitted predominantly commercial solutions, whereas R&D professionals favored technical approaches. When teams worked without AI, they produced more balanced solutions through cross-functional collaboration. Interestingly, individuals using AI achieved similar levels of solution balance on their own, effectively replicating the knowledge integration typically achieved through team collaboration. This suggests AI serves not just as an information provider but as an effective boundary-spanning mechanism, helping professionals reason across traditional domain boundaries and approach problems more holistically.

Complementing the expertise shift, our experimental design seems to reveal a well-documented economic principle at work: diminishing marginal returns in team expansion. As we view GenAI as a cybernetic teammate, we can conceptualize our experimental

conditions onto a sequential expansion of team head count. We move from individual to dyad (human–human or human–AI) to triad (human–human–AI), and our results suggest that the first teammate addition, regardless of type, delivered significant average quality gains.²⁴ However, the subsequent addition yielded less significant average improvement, while increasing the likelihood of producing top-decile, breakthrough ideas. This pattern echoes classic organizational research findings where additional team members simultaneously contribute knowledge while increasing coordination complexity and diffusing individual accountability (Steiner 1972, Latané et al. 1979, Hambrick and D'Aveni 1992, Bernerth et al. 2023). Importantly, the cross-functional composition of commercial and R&D expertise in our dyads already matches the professional diversity that P&G's experience suggests is most critical for early-stage product development. This pattern may indicate that optimal team configurations depend more on capturing essential functional expertise than on raw head count.

Although our design does not permit a clean separation between AI augmenting noncollaborative tasks and AI substituting for collaborative functions, the magnitudes in Table 2 offer an illustrative guide. Individuals using AI performed about 0.37 standard deviations higher than individuals without AI, whereas teams using AI outperformed teams without AI by about 0.15 standard deviations. This pattern suggests that roughly 40% of the solo performance improvement likely reflects AI enhancing noncollaborative aspects of work that also benefit teams, whereas the remaining 60% appears linked to AI substituting for some collaborative functions that human teammates typically provide. Because individuals and teams engage with AI in systematically different ways, this comparison should be viewed as indicative rather than causal: it captures overlapping but not identical processes of AI use across contexts. We view this as an indicative rather than definitive decomposition, but it highlights that a meaningful portion of AI's value for solo workers comes from its capacity to partially replicate the benefits of teamwork, rather than from surface-level assistance alone.

Another important benefit of teamwork is the emotional boost it provides all along the process. In that regard, our result on the positive impact of AI on workers' experience are particularly noteworthy. Contrary to fears about AI creating negative workplace experiences, we found consistently positive emotional responses to AI use, including increased excitement and enthusiasm, as well as reduced anxiety and frustration. Unlike some earlier waves of technological change, and even earlier iterations of AI technologies (Stein et al. 2015, Glikson and Woolley 2020, Dell'Acqua et al. 2025), GenAI's interactive features appear to create positive experiences for workers, aligning with

emerging evidence on the psychological effects of conversational AI (Li et al. 2023, De Freitas et al. 2024, Riedl and Weidmann 2025). These findings should be interpreted cautiously, as they reflect immediate reactions to AI collaboration rather than the complex social dynamics that develop in long-term team relationships. Moreover, although AI reduced negative emotions and friction in our setting, some level of constructive disagreement can be valuable for creative exploration, meaning that lower tension does not always imply better collaboration.

These emotional patterns, in turn, suggest implications for learning dynamics in AI-augmented work environments. The positive emotional responses to AI collaboration could create a self-reinforcing cycle if participants who reported the most positive emotional experiences while working independently with AI showed the strongest preference for AI collaboration over human teammates. The results of Tables 7 and 8 suggest that positive initial AI experiences are correlated with a higher expected likelihood of AI use in the participants' future work. Hence, positive affect may help turn a one-shot productivity boost into more sustained adoption, potentially accelerating organizational learning curves.

Beyond examining whether AI improves outcomes, our findings also help clarify how and where it contributes to the creative process. Two common questions arise: Does AI merely improve the presentation of ideas rather than their substantive quality? And does AI homogenize the quality distribution? First, regarding presentation versus substance, our evidence suggests that AI improves idea quality. We find no relationship between solution length and quality ratings among non-AI participants, indicating evaluators did not simply reward longer or more polished text. Similarly, controlling for typographical errors does not meaningfully alter our treatment effects.²⁵ Our decomposition analysis further suggests that AI's benefits emerge at the idea-generation stage itself: AI significantly increases the average quality of the five initial ideas participants generate before selecting one to develop further (Figure 8, panel (a)), suggesting the improvements reflect better concepts from the outset rather than superior polish. Second, regarding homogenization, we find that AI elevates rather than compresses the variance of quality distribution. The range between the highest- and lowest-quality ideas remains similar across all conditions, suggesting AI elevates the entire quality distribution rather than narrowing it toward a homogeneous mean.²⁶ Moreover, AI-augmented teams were significantly more likely to produce high-quality ideas in the top decile of the distribution, with Team + AI showing roughly three times the likelihood of breakthrough solutions compared with individuals without AI, indicating

that AI enhances rather than diminishes innovation potential.

Taken together, these results indicate that AI may be more than a passive tool and may function as a cybernetic teammate. By interfacing with human problem solvers—providing real-time feedback, stimulating the ideation process, bridging cross-functional expertise, and influencing self-reported emotional states—GenAI appears able to perform roles we typically associate with human collaborators. In this sense, AI not only enhances individual cognitive work but may also perform some collective functions, such as ideation and iterative refinement, helping teams address complex challenges. Although AI cannot fully replicate the richness of human social and emotional interaction, its ability to contribute to collaborative work suggests the possibility of changes in how knowledge work is structured and carried out.²⁷

This view aligns with a body of literature that conceptualizes AI not merely as a tool or a medium, but as an active “counterpart” within broader sociotechnical systems. Drawing on distributed cognition (Hutchins 1991, 1995) and actor–network theory (Callon 1984; Latour 1987, 2007), recent organizational scholarship argues for examining AI as an active counterpart in systems of work involving multiple organizational actors and technologies (Anthony et al. 2023). Our study supports and extends these arguments by suggesting that GenAI can shape expertise sharing, team dynamics, and social engagement in ways that extend beyond the traditional boundaries of automation. In other words, AI’s role may be more than that of a tool or facilitator, affecting patterns of collaboration. By treating AI as an active counterpart, we gain insight into how GenAI mediates, and is mediated by, the collective processes involved in teamwork.

Along these lines, conceptualizing AI as a cybernetic teammate raises important questions about how humans develop and apply theories of mind in human–AI collaboration (Kelley 1973, Malle 2006). Just as effective human teamwork relies on understanding teammates’ cognitive processes, motivations, and decision-making patterns, successful collaboration with AI will lead humans to form (possibly inaccurate) mental models of how AI systems process information, generate responses, and approach problems (Glikson and Woolley 2020, Lebovitz et al. 2022, Anthony et al. 2023). These “theories of the AI mind” may significantly influence collaboration effectiveness. Unpacking how such theories arise, and whether they track the technology’s jagged capabilities (Dell’Acqua et al. 2023) remains a critical frontier for effective human–AI collaboration.

These findings have significant organizational implications. First, organizations may need to reevaluate optimal team sizes and compositions. The fact that AI-enabled individuals can perform at levels

comparable to traditional teams suggests opportunities for more flexible and efficient organizational structures. At the same time, an important nuance emerges when considering top-tier solutions: AI-augmented teams were more likely to produce proposals ranking in the top decile, underscoring the unique synergy produced by combining human collaboration with AI-based augmentation. This may be a crucial consideration for organizations, as different firms may respond differently. Some firms may focus on the efficiency side, whereas others may focus on the complementarity.²⁸ The increased quality and comprehensiveness of AI-enabled work suggest opportunities to redesign work processes and deliverable expectations. Organizations should invest in developing their workers’ AI interaction capabilities, as this appears to be an increasingly critical skill. Given AI’s ability to break down silos, there may also be value in training workers to think more broadly across functional boundaries.

Two important caveats shape the interpretation of these findings. First, our participants were relatively inexperienced with AI prompting techniques, suggesting the observed benefits may represent a lower bound. As users develop more sophisticated AI interaction strategies, the advantages of AI-enabled work could increase substantially. Second, the AI tools used were not optimized for collaborative work environments. Purpose-built collaborative AI systems could potentially unlock significantly greater benefits by better supporting group dynamics and collective problem-solving processes. Related to this, we should also highlight two organizational limitations. First, although we followed the firm’s early-stage product development routine, our experiment relied on one-day virtual collaborations that did not fully capture the day-to-day complexities of team interactions in organizations—such as extended coordination challenges and iterative rework cycles. Second, we focused on cross-functional pairs of human workers, whereas collaborations involving team members with similar expertise, or in larger, more intricate team structures, may exhibit different patterns of AI adoption and effectiveness.

Several scope conditions further shape the generalizability of our findings. Our study was conducted within a single company in the consumer goods industry, focusing on early-stage new product development through virtual interactions between largely unfamiliar participants. These conditions resemble flash teams rather than established organizational teams with embedded relationships and knowledge (Retelny et al. 2014, Valentine and Edmondson 2015, Valentine and Bernstein 2025). Additionally, our findings reflect the capabilities of a single AI model at a particular point in time, and all collaborations occurred remotely, where the dynamics of human–AI collaboration may differ from face-to-face

settings that involve nonverbal communication and physical presence.

7. Conclusion

Our research suggests that AI adoption may require reconsidering assumptions about team structures and organizational design. By showing that AI can raise individual performance to levels comparable to traditional teams while also reducing professional silos, our findings contribute to both the emerging literature on AI in organizations and classical theories of team effectiveness. The increased likelihood of exceptional performance in AI-enabled teams, combined with evidence of reduced functional boundaries and positive emotional effects, suggests interactions between human and artificial capabilities that merit further investigation. As organizations continue to integrate AI technologies, understanding these dynamics may be important for organizational theory and practice. Future research should examine how these patterns evolve as users develop greater AI proficiency, how different organizational contexts moderate these effects, and how sustained AI use impacts the development and transfer of expertise within organizations.

Against this backdrop, our findings suggest several promising avenues for future research. First, how do the benefits of AI integration evolve as users become more sophisticated in their AI interactions? Given our participants' relative inexperience with AI, understanding the learning curve and potential ceiling effects becomes crucial. Second, researchers should investigate the economic principles governing team composition in AI-augmented environments. Our observation of diminishing returns when expanding from dyads to triads raises several questions about optimal team sizing in the presence of AI teammates. Future work could systematically vary both skill complementarity and human–AI ratios to determine whether AI fundamentally alters traditional team-scaling principles. Finally, given the rapid pace of AI model advancement and increasing use of AI in organizations (Bick et al. 2026), it is incumbent upon researchers to consider conducting “clinical trials” of AI in partnership with organizations. We believe that management scholars can have a significant say in the rate and direction of AI adoption and usage inside of organizations if they can marshal causal evidence on AI's positive and negative impact on individuals, teams and organizations.

Furthermore, our findings on breakthrough innovations warrant deeper examination of exploration–exploitation dynamics in AI collaboration. Does AI's ability to enhance the right tail of performance apply across different innovation contexts and task complexities? This connects to multiple questions about expertise development, especially given recent findings on how

expertise remains key to leveraging the use of technology for innovative processes (Lazar et al. 2025): How does AI integration affect the development of domain expertise over time? What features of AI systems specifically support effective knowledge integration across professional boundaries? Does AI-enabled boundary spanning foster genuine knowledge growth, or merely facilitate temporary access to existing expertise? Studies examining how positive initial AI experiences might create self-reinforcing adoption cycles could also clarify the emotional path of AI interactions within organizations.

Overall, our findings suggest that AI may be more than an advanced search engine or text generator, instead playing a more active role in collaborative work. By contributing to decision making, creativity, and emotional responses, AI may affect the conditions under which teams form and function. Although questions remain about how AI will influence long-term skill development and trust, our evidence points to the possibility of broader changes in knowledge work, raising new questions about the evolving interplay between human and machine contributions.

Acknowledgments

The authors thank Ramona Pop for her critical help managing the experiment. The authors thank Andrea Dorbu, Bandy Chin, Corey Gelb-Bicknell, Conor Mackey, Hadi Abbas, John Kalil, Michael Menietti, Sarah Stegall-Rodriguez, and Vishnu Kulkarni for very helpful support and research assistance. The authors thank Iavor Bojinov, Jacqueline Lane, Simon Friis, and Brent Hecht for thoughtful comments. Seminar participants at Harvard, New York University, ESSEC Business School, INSEAD, Bristol University, London Business School, Michigan Ross, Berkeley, Northeastern, Stanford, Wharton, OpenAI, the Organisation for Economic Co-operation and Development, the French Department of the Treasury, and the European Commission provided helpful feedback. The authors are grateful for the guidance of Sharique Hasan and three referees at *Organization Science* in helping to improve this paper. Procter & Gamble provided financial support to the HBS AI Institute through gifts to Harvard Business School during the period 2023–2025. Karim Lakhani received compensation as a consultant for Procter & Gamble for the period 2021–2022. The authors maintained full intellectual independence throughout the study. Author contributions are as follows: F.D.A., C.A., R.S., and K.R.L. established the collaboration and designed the experiment. F.D.A., C.A., and K.R.L. oversaw the execution of the experiment, and led the analyses, framing, and write-up. H.L. and E.M. contributed to the experimental design. H.L., R.S., and E.M. contributed to the framing. E.M. and L.M. contributed to the prompt training. Y.H., J.G., H.N., and S.T. facilitated organizational access and enabled the execution of the experiment inside P&G. All coauthors contributed to revisions of the manuscript. They used Claude, Manus, and ChatGPT for light copyediting.

Endnotes

¹ Bernerth et al. (2023, pp. 1230–1231) define workplace teams as “two or more individuals who share some interdependency and responsibility of work tasks and collective outputs.”

² The term draws from Norbert Wiener’s (1948, 1950) foundational work on cybernetics, which describes feedback-regulated systems that dynamically adjust their behavior in response to environmental inputs. Rather than simply automating tasks, such systems modify their functioning through iterative feedback loops, a property that makes them capable of participating in collaborative processes.

³ This builds on existing literature investigating the adoption and impact of earlier waves of AI technologies. See, for example, Brynjolfsson et al. (2019, 2018), Agrawal et al. (2018), Furman and Seamans (2019), Iansiti and Lakhani (2020), Raisch and Krakowski (2021), Jacobides et al. (2021), and McElheran et al. (2024).

⁴ This project (IRB24-0202) received institutional review board approval.

⁵ The study was preregistered at the American Economic Association Randomized Controlled Trial Registry (AEARCTR-0013603), detailing our experimental conditions, outcome variables, and analytical approaches.

⁶ A long literature in management confirms the benefit of this approach for successful innovation (e.g., Dougherty 1992).

⁷ The detailed description of the tasks given to participants can be found in Online Appendix A.

⁸ Of the 791 participants who completed the workshop, 776 provided complete data including post-task surveys. The 15 participants with incomplete surveys were distributed across conditions (4 from non-AI teams and 11 from AI teams), and their exclusion does not affect our main results.

⁹ Thirty-five participants were not randomly assigned either because they entered the product development workshop too late (in which case they completed the task alone without AI) or because their seniority was above band 3 (in which case they completed the task alone with AI). All our analyses exclude these participants but our results are consistent when we include these nonrandomized participants. These participants were isolated in separate sessions and did not interact with randomly assigned participants, eliminating potential contamination effects.

¹⁰ The randomization clusters included a geographical component primarily in order to accommodate time-zone differences and ensure that team members could collaborate in real time.

¹¹ The nested structure refers to individuals being grouped within teams, which are further nested within business units and geographical regions, requiring careful statistical consideration. Maintaining team integrity posed a significant challenge; if one member of a two-person team failed to participate, the entire team was nullified, leading us to automatically reassign individuals from incomplete teams to individual assignments to preserve data collection opportunities.

¹² The random assignment of leadership role between R&D and commercial professionals had no statistically significant impact on any of our team outcomes.

¹³ Participants at the July workshop had access to GPT-4o. The results remain consistent across the various workshop sessions.

¹⁴ Interestingly, despite providing structured prompts to all participants in AI-enabled conditions, only 38% actually utilized these suggested prompts in their interactions. Participants who did not follow the prompt guidance achieved performance levels equivalent to those who did, with both groups significantly outperforming the control conditions. This pattern suggests that participants quickly adapted the AI tool to their own working styles rather than relying on prescribed approaches, and that AI’s benefits reflect authentic,

self-directed collaboration rather than dependence on specific prompting techniques.

¹⁵ LLM capabilities are rapidly evolving, and our specific effect sizes should be interpreted as directional indicators rather than precise estimates that will hold across all future models. However, the core mechanisms we identify (AI’s ability to provide continuous creative input, reduce ideation fatigue, and substitute for certain collaborative functions) represent core capabilities that are likely to persist and strengthen as LLMs improve (Xiao et al. 2025) and get increasingly adopted by organizations (Bick et al. 2026).

¹⁶ Note that we consider the average of these values for the team conditions (with and without AI). Results are robust to the use of alternative specifications for these controls such as the sum or the minimum or maximum of the team value.

¹⁷ As a robustness check, we also estimate all models with standard errors clustered at the randomization-unit level (eight clusters defined by business unit \times geography) using wild cluster bootstrap for clustered regressions. Results remain substantively unchanged (see Table A3 in the Online Appendix).

¹⁸ Evaluators assessed five dimensions of each solution, *Quality*, *Novelty*, *Feasibility*, *Impact*, and *Business Potential*, using the same 1–10 scale and blinded evaluation process. When we combine these four measures into a composite index of overall quality, our results replicate. When we disaggregate them, we find that *Novelty*, *Impact*, and *Business Potential* closely mirror the patterns reported for overall *Quality*, whereas no significant differences emerge for *Feasibility*.

¹⁹ As a robustness check, we replicated all analyses using AI-generated evaluations of the solutions. Results remain consistent across all models.

²⁰ For two-person teams, we construct the team-level outcome by averaging the individual participants’ post-task changes in these composite measures.

²¹ Positive and negative emotions show no significant differences between conditions in the preexperimental period, as can be seen in Table 1.

²² Teams where only one employee has as their core job to work on new product development are classified as core-job teams. For teams without AI, teams with one core-job participant are indistinguishable from teams composed of two core-job participants.

²³ Although we cannot directly observe downstream development or commercialization decisions, this early-stage product development task represents a core component of P&G’s innovation pipeline. Our workshop involved approximately 800 professionals; running a session of this scale would represent an investment exceeding one million dollars. In this context, even a modest increase in the probability of advancing a high-impact product concept could translate into hundreds of millions of dollars in additional expected revenue, underscoring the economic significance of the productivity gains we document.

²⁴ Although the point estimates are directionally consistent with diminishing marginal returns, the incremental differences between conditions are not statistically significant.

²⁵ See Table A2 in the Online Appendix.

²⁶ We do observe, however, that AI-assisted solutions exhibit greater semantic similarity to one another in embedding space, consistent with recent evidence on LLM-driven content homogenization (Doshi and Hauser 2024, Wang et al. 2026). Whether this semantic convergence carries implications for organizational innovation diversity over time is an important question for future research.

²⁷ See Leonardi and Neeley (2022) and Farrell et al. (2025) for related discussions.

²⁸ Our partner P&G was squarely focused on the potential for top quality solutions.

References

- Agrawal A, Gans J, Goldfarb A (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press, Boston).
- Alchian AA, Demsetz H (1972) Production, information costs, and economic organization. *Amer. Econom. Rev.* 62(5):777–795.
- Ancona DG, Caldwell DF (1992) Bridging the boundary: External activity and performance in organizational teams. *Admin. Sci. Quart.* 37(4):634–665.
- Anthony C, Bechky BA, Fayard AL (2023) “Collaborating” with AI: Taking a system view to explore the future of work. *Organ. Sci.* 34(5):1672–1694.
- Argote L (1999) *Organizational Learning: Creating, Retaining and Transferring Knowledge* (Kluwer Academic Publishers, Norwell, MA).
- Argote L, Lee S, Park J (2021) Organizational learning processes and outcomes: Major findings and future research directions. *Management Sci.* 67(9):5399–5429.
- Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, et al. (2023) Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine* 183(6):589–596.
- Ayoubi C, Lane JN, Szajnarfarber Z (2026) The two faces of expertise: How skills and experience shape the evaluation of innovation. Working paper, ESSEC Business School, Cergy, France.
- Ayoubi C, Pezzoni M, Visentin F (2017) At the origins of learning: Absorbing knowledge flows from within the team. *J. Econom. Behav. Organ.* 134:374–387.
- Bailey DE, Leonardi PM, Chong J (2010) Minding the gaps: Understanding technology interdependence and coordination in knowledge work. *Organ. Sci.* 21(3):713–730.
- Balasubramanian N, Ye Y, Xu M (2022) Substituting human decision-making with machine learning: Implications for organizational learning. *Acad. Management Rev.* 47(3):448–465.
- Beane M (2019) Shadow learning: Building robotic surgical skill when approved means fail. *Admin. Sci. Quart.* 64(1):87–123.
- Beane M, Anthony C (2024) Inverted apprenticeship: How senior occupational members develop practical expertise and preserve their position when new technologies arrive. *Organ. Sci.* 35(2):405–431.
- Beaudry A, Pinsonneault A (2010) The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. *MIS Quart.* 34(4):689–710.
- Becker GS, Murphy KM (1992) The division of labor, coordination costs, and knowledge. *Quart. J. Econom.* 107(4):1137–1160.
- Bernerth JB, Beus JM, Helmuth CA, Boyd TL (2023) The more the merrier or too many cooks spoil the pot? A meta-analytic examination of team size and team effectiveness. *J. Organ. Behav.* 44(8):1230–1262.
- Bick A, Blandin A, Deming DJ (2026) The rapid adoption of generative AI. *Management Sci.*, ePub ahead of print January 20, <https://doi.org/10.1287/mnsc.2025.02523>.
- Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Sci.* 57(5):843–863.
- Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Sci.* 62(10):2765–2783.
- Boussioux L, Lane JN, Zhang M, Jacimovic V, Lakhani KR (2025) The crowdless future? How generative AI is shaping the future of human crowdsourcing. *Organ. Sci.* 35(5):1589–1607.
- Brown SL, Eisenhardt KM (1995) Product development: Past research, present findings, and future directions. *Acad. Management Rev.* 20(2):343–378.
- Brynjolfsson E, Li D, Raymond LR (2025) Generative AI at work. *Quart. J. Econom.* 140(2):889–942.
- Brynjolfsson E, Mitchell T, Rock D (2018) What can machines learn and what does it mean for occupations and the economy? *AEA Papers Proc.* 108:43–47.
- Brynjolfsson E, Rock D, Syverson C (2019) Artificial intelligence and the modern productivity paradox. Agrawal A, Gans J, Goldfarb A, eds. *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Chicago), 23–57.
- Brynjolfsson E, Rock D, Syverson C (2021) The productivity J-curve: How intangibles complement general purpose technologies. *Amer. Econom. J.: Macroeconomics* 1(13):333–372.
- Callon M (1984) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. Law J, ed. *Power, Action and Belief: A New Sociology of Knowledge?* (Routledge, Boston), 196–223.
- Cattani G, Ferriani S, Lanza A (2017) Deconstructing the outsider puzzle: The legitimization journey of novelty. *Organ. Sci.* 28(6):965–992.
- Choudhary V, Marchetti A, Shrestha YR, Puranam P (2025) Human-AI ensembles: When can they work? *J. Management* 51(2):536–569.
- Cohen SG, Bailey DE (1997) What makes teams work: Group effectiveness research from the shop floor to the executive suite. *J. Management* 23(3):239–290.
- Csaszar FA (2012) Organizational structure as a determinant of performance: Evidence from mutual funds. *Strategic Management J.* 33(6):611–632.
- Csaszar FA, Ketkar H, Kim H (2024) Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Sci.* 9(4):322–345.
- Dahan E, Mendelson H (2001) An extreme-value model of concept testing. *Management Sci.* 47(1):102–116.
- De Freitas JD, Uguralp AK, Uguralp Z, Puntoni S (2024) AI companions reduce loneliness. Preprint, submitted July 26, <https://doi.org/10.2139/ssrn.4893097>.
- Dell'Acqua F (2022) Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters. Working paper, Harvard Business School, Boston.
- Dell'Acqua F, Kogut B, Perkowski P (2025) Super Mario meets AI: Experimental effects of automation and skills on team performance and coordination. *Rev. Econom. Statist.* 107(4):951–966.
- Dell'Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Kraye L, Candelon F, Lakhani KR (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Working Paper 24-013, Technology and Operations Management Unit, Harvard Business School, Boston.
- Deming DJ (2017) The growing importance of social skills in the labor market. *Quart. J. Econom.* 132(4):1593–1640.
- Deutsch M (1949) A theory of co-operation and competition. *Human Relations* 2(2):129–152.
- DiBenigno J, Kellogg KC (2014) Beyond occupational differences: The importance of cross-cutting demographics and dyadic toolkits for collaboration in a US hospital. *Admin. Sci. Quart.* 59(3):375–408.
- Doshi AR, Hauser OP (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci. Adv.* 10(28):eadn5290.
- Doshi AR, Bell JJ, Mirzayev E, Vanneste BS (2025) Generative artificial intelligence and evaluating strategic decisions. *Strategic Management J.* 46(3):583–610.
- Dougherty D (1992) Interpretive barriers to successful product innovation in large firms. *Organ. Sci.* 3(2):179–202.
- Eloundou T, Manning S, Mishkin P, Rock D (2024) GPTs are GPTs: An early look at the labor market impact potential of large language models. *Science* 384(6702):1306–1308.
- Faraj S, Sproull L (2000) Coordinating expertise in software development teams. *Management Sci.* 46(12):1554–1568.

- Faraj S, Pachidi S, Sayegh K (2018) Working and organizing in the age of the learning algorithm. *Inform. Organ.* 28(1):62–70.
- Farrell H, Gopnik A, Shalizi C, Evans J (2025) Large AI models are cultural and social technologies. *Science* 387(6739):1153–1156.
- Furman J, Seamans R (2019) AI and the economy. *Innovation Policy Econom.* 19(1):161–191.
- Garud R (1997) On the distinction between know-how, know-why, and know-what. *Adv. Strategic Management* 14:81–101.
- Girotra K, Terwiesch C, Ulrich KT (2010) Idea generation and the quality of the best idea. *Management Sci.* 56(4):591–605.
- Glikson E, Woolley AW (2020) Human trust in artificial intelligence: Review of empirical research. *Acad. Management Ann.* 14(2):627–660.
- Hambrick D, D'Aveni R (1992) Top team deterioration as part of downward spiral of large corporate bankruptcies. *Management Sci.* 38(10):1445–1466.
- Hoffmann M, Boysel S, Nagle F, Peng S, Xu K (2024) Generative AI and the nature of work. CESifo working paper, Center for Economic Studies, Ludwig-Maximilians-Universität München, Munich, Germany.
- Hutchins E (1991) Organizing work by adaptation. *Organ. Sci.* 2(1):14–39.
- Hutchins E (1995) *Cognition in the Wild* (MIT Press, Cambridge, MA).
- Iansiti M, Lakhani KR (2020) *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World* (Harvard Business Review Press, Boston).
- Jacobides MG, Brusoni S, Candelon F (2021) The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Sci.* 6(4):412–435.
- Jeppesen LB, Lakhani KR (2010) Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* 21(5):1016–1033.
- Johnson DW, Johnson RT (2005) New developments in social interdependence theory. *Genetic Soc. General Psych. Monographs* 131(4):285–358.
- Jonassen Z, He VF, von Krogh G (2026) Good lessons despite bad feelings: How boundary-spanning teams learn from collaboration failure. *Organ. Sci.* 37(1):17–47.
- Jones BF (2009) The burden of knowledge and the “death of the Renaissance man”: Is innovation getting harder? *Rev. Econom. Stud.* 76(1):283–317.
- Kacperczyk A, Younkin P (2017) The paradox of breadth: The tension between experience and legitimacy in the transition to entrepreneurship. *Admin. Sci. Quart.* 62(4):731–764.
- Kelley HH (1973) The processes of causal attribution. *Amer. Psych.* 28(2):107–128.
- Kellogg KC, Orlikowski WJ, Yates J (2006) Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organ. Sci.* 17(1):22–44.
- Kellogg KC, Valentine MA, Christin A (2020) Algorithms at work: The new contested terrain of control. *Acad. Management Ann.* 14(1):366–410.
- Kogut B, Zander U (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* 3(3):383–397.
- Kozlowski SWJ, Bell BS (2013) Work groups and teams in organizations: Review update. Schmitt N, Highhouse S, eds. *Handbook of Psychology, Vol. 12: Industrial and Organizational Psychology*, 2nd ed. (Wiley, Hoboken, NJ), 412–469.
- Lane JN (2023) The subjective expected utility approach and a framework for defining project risk in terms of novelty and feasibility—A response to Franzoni and Stephan (2023), “uncertainty and risk-taking in science.” *Res. Policy* 52(3):104707.
- Lane JN, Boussioux L, Ayoubi C, Hao Chen Y, Lin C, Spens R, Wagh P, Wang PH (2026) The narrative AI advantage? A field experiment on AI-augmented evaluations of early-stage innovations. Working Paper No. 25-001, Harvard Business School, Boston.
- Latané B, Williams K, Harkins S (1979) Many hands make light the work: The causes and consequences of social loafing. *J. Personality Soc. Psych.* 37(6):822–832.
- Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society* (Harvard University Press, Cambridge, MA).
- Latour B (2007) *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford University Press, Oxford, UK).
- Lazar M, Lifshitz H, Ayoubi C, Emuna H (2025) Would Archimedes shout “eureka” with algorithms? The hidden hand of algorithmic design in idea generation, the creation of ideation bubbles, and how experts can burst them. *Acad. Management J.* 68(5):881–906.
- Lazer D, Katz N (2003) Building effective intra-organizational networks: The role of teams. Working paper, Northeastern University, Boston.
- Lebovitz S, Lifshitz-Assaf H, Levina N (2022) To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organ. Sci.* 33(1):126–148.
- Leonardi P, Neeley T (2022) *The Digital Mindset: What It Really Takes to Thrive in the Age of Data, Algorithms, and AI* (Harvard Business Review Press, Boston).
- Levina N, Vaast E (2005) The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quart.* 29(2):335–363.
- Levinthal DA (1997) Adaptation on rugged landscapes. *Management Sci.* 43(7):934–950.
- Li JZ, Herderich A, Goldenberg A (2024) Skill but not effort drive GPT overperformance over humans in cognitive reframing of negative scenarios. Working paper, Harvard University, Cambridge, MA.
- Li D, Raymond LR, Bergman P (2026) Hiring as exploration. *Rev. Econom. Stud.* 93(2):1200–1240.
- Li H, Zhang R, Lee Y-C, Kraut RE, Mohr DC (2023) Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine* 6(1):236.
- Lindbeck A, Snower DJ (2000) Multitask learning and the reorganization of work: From Tayloristic to holistic organization. *J. Labor Econom.* 18(3):353–376.
- Malle BF (2006) *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction* (MIT Press, Cambridge, MA).
- March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.
- March JG, Simon HA (1958) *Organizations* (John Wiley & Sons, New York).
- McElheran K, Li JF, Brynjolfsson E, Kroff Z, Dinlersoz E, Foster L, Zolas N (2024) AI adoption in America: Who, what, and where. *J. Econom. Management Strategy* 33(2):375–415.
- Mollick E (2024) *Co-Intelligence* (Random House, London).
- Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change* (Harvard University Press, Cambridge, MA).
- Nickerson JA, Zenger TR (2004) A knowledge-based theory of the firm—The problem-solving perspective. *Organ. Sci.* 15(6):617–632.
- Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381(6654):187–192.
- Orlikowski WJ (2002) Knowing in practice: Enacting a collective capability in distributed organizing. *Organ. Sci.* 13(3):249–273.
- Otis N, Clarke RP, Delecourt S, Holtz D, Koning R (2024) The uneven impact of generative AI on entrepreneurial performance. Preprint, submitted January 17, <https://doi.org/10.2139/ssrn.4671369>.
- Page SE (2019) *The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy* (Princeton University Press, Princeton, NJ).
- Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of AI on developer productivity: Evidence from GitHub copilot. Preprint, submitted February 13, <https://arxiv.org/abs/2302.06590>.
- Puranam P (2018) *The Microstructure of Organizations* (Oxford University Press, New York).

- Raisch S, Fomina K (2025) Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Acad. Management Rev.* 50(2):441–464.
- Raisch S, Krakowski S (2021) Artificial intelligence and management: The automation–augmentation paradox. *Acad. Management Rev.* 46(1):192–210.
- Raj M, Seamans R (2019) Primer on artificial intelligence and robotics. *J. Organ. Design.* 8(1):11.
- Randazzo S, Joshi A, Kellogg KC, Lifshitz H, Dell'Acqua F, Lakhani KR (2025a) GenAI as a power persuader: How professionals get persuasion bombed when they attempt to validate LLMs. Working paper, Warwick Business School, Coventry, UK.
- Randazzo S, Lifshitz-Assaf H, Kellogg K, Dell'Acqua F, Mollick ER, Candelon F, Lakhani KR (2025b) Cyborgs, centaurs and self-automators: The three modes of human-GenAI knowledge work and their implications for skilling and the future of expertise. The Wharton School Research Paper, Harvard Business School Working Paper, (26-036), 26-036.
- Retelny D, Robaszkievicz S, To A, Lasecki WS, Patel J, Rahmati N, Doshi T, Valentine M, Bernstein MS (2014) Expert crowdsourcing with flash teams. *Proc. 27th Annual ACM Sympos. User Interface Software Tech.* (Association for Computing Machinery, New York), 75–85.
- Riedl C, Weidmann B (2025) Quantifying human-AI synergy. Working paper.
- Rivkin JW (2000) Imitation of complex strategies. *Management Sci.* 46(6):824–844.
- Singh J, Fleming L (2010) Lone inventors as sources of breakthroughs: Myth or reality? *Management Sci.* 56(1):41–56.
- Souitaris V, Peng B, Zerbinati S, Shepherd DA (2023) Specialists, generalists, or both? Founders' multidimensional breadth of experience and entrepreneurial ventures' fundraising at IPO. *Organ. Sci.* 34(2):557–588.
- Stein MK, Newell S, Wagner EL, Galliers RD (2015) Coping with information technology. *MIS Quart.* 39(2):367–392.
- Steiner ID (1972) *Group Process and Productivity* (Academic Press, New York).
- Tarafdar M, Cooper CL, Stich JF (2019) The technostress trifecta—Techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Inform. Systems J.* 29(1):6–42.
- Teodoridis F (2018) Understanding team knowledge production: The interrelated roles of technology and expertise. *Management Sci.* 64(8):3625–3648.
- Terwiesch C, Loch CH (2004) Collaborative prototyping and the pricing of custom-designed products. *Management Sci.* 50(2):145–158.
- Terwiesch C, Xu Y (2008) Innovation contests, open innovation, and multiagent problem solving. *Management Sci.* 54(9):1529–1543.
- Valentine M, Bernstein M (2025) *Flash Teams: Leading the Future of AI-Enhanced, On-Demand Work* (MIT Press, Cambridge, MA).
- Valentine MA, Edmondson AC (2015) Team scaffolds: How mesolevel structures enable role-based coordination in temporary groups. *Organ. Sci.* 26(2):405–422.
- Vuori TO, Huy QN (2016) Distributed attention and shared emotions in the innovation process: How Nokia lost the smartphone battle. *Admin. Sci. Quart.* 61(1):9–51.
- Wang D, Huang D, Shen H, Uzzi B (2026) A large-scale comparison of divergent creativity in humans and large language models. *Nature Human Behav.* 10:531–540.
- Weber RA, Camerer CF (2003) Cultural conflict and merger failure: An experimental approach. *Management Sci.* 49(4):400–415.
- Weidmann B, Deming DJ (2020) Team players: How social skills improve group performance. NBER Working Paper No. 27071, National Bureau of Economic Research, Cambridge, MA.
- Wiener N (1948) *Cybernetics: Or Control and Communication in the Animal and the Machine* (MIT Press, Cambridge, MA).
- Wiener N (1950) *The Human Use of Human Beings: Cybernetics and Society*, 1st ed. (Houghton Mifflin, Boston).
- Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036–1039.
- Xiao C, Cai J, Zhao W, Lin B, Zeng G, Zhou J, Zheng Z, Han X, Liu Z, Sun M (2025) Densing law of LLMs. *Nature Machine Intelligence* 7:1823–1833.
- Zander U, Kogut B (1995) Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organ. Sci.* 6(1):76–92.
-
- Fabrizio Dell'Acqua** is a postdoctoral researcher at Harvard Business School and HBS AI Institute. He received his PhD in management from Columbia Business School. His research examines how human-AI collaboration reshapes knowledge work at the individual, team, and organizational levels. Prior to his PhD, he received degrees in economics from Bocconi University and London Business School.
- Charles Ayoubi** is an assistant professor of management at ESSEC Business School, Paris, France. He received his PhD in innovation economics from the École Polytechnique Fédérale de Lausanne. Prior to joining ESSEC, he was a postdoctoral research fellow at Harvard Business School in the Digital Data Design Institute. His research explores how organizations generate, evaluate, and diffuse innovative ideas, with a focus on how generative AI is reshaping decision making, and business opportunities.
- Hila Lifshitz** is a professor of management at Warwick Business School and affiliated faculty at Harvard's Digital Data Design Institute. She is the head of the Artificial Intelligence Innovation Network at Warwick University. She conducts field studies exploring the transformation of day-to-day knowledge work processes and the use of AI for innovation processes as well as for critical decision-making processes. She earned her doctorate from Harvard Business School.
- Raffaella Sadun** is the Charles E. Wilson Professor of Business Administration at Harvard Business School (HBS). She received her PhD in economics from the London School of Economics. Her research focuses on managerial and organizational drivers of productivity and growth, with emphasis on the measurement of management practices across organizations and countries. She cofounded the World Management Survey and coleads the Digital Reskilling Lab at HBS.
- Ethan Mollick** is the Ralph J. Roberts Distinguished Faculty Scholar, a Rowan Fellow, and an associate professor of management at the Wharton School of the University of Pennsylvania. He received his PhD and MBA from the Massachusetts Institute of Technology Sloan School of Management. His research interests include the effects of artificial intelligence on work and education, with particular emphasis on how emerging technologies transform organizational processes and individual performance.
- Lilach Mollick** is the codirector of the Wharton Generative AI Labs. Her work focuses on the development of pedagogical strategies that include artificial intelligence and interactive methodologies. She has worked with Wharton to develop a wide range of educational tools and games used in classrooms worldwide. She has also written several papers on the uses of AI for teaching and training, and her work on AI has been discussed in publications including the *New York Times* and *Vox*.
- Yi Han** is a researcher and innovation leader at Procter & Gamble. His work focuses on the application of artificial intelligence and digital technologies in innovation processes, with particular emphasis on how large organizations integrate emerging technologies to drive product and business transformation. His research interests include AI-enabled innovation, organizational capabilities, and digital transformation.
- Jeff Goldman** is vice president of enterprise AI at P&G, leading P&G's global AI organization across data science, AI engineering,

and AI factory. He founded P&G's Global Data Science organization, served as analytic advisor to P&G's C-suite, led business analytics for Global Markets and the Western European Analytics, and founded the Business Analytics group for China and Product Supply Analytics for Asia. He holds a BA in economics and a master's in operations research from Cornell.

Hari Nair is vice president of R&D at Procter & Gamble. He received his BSc in chemical engineering from University of Wisconsin–Madison and is also a 2019 Harvard Advanced Leadership Initiative Fellow.

Stew Taub is vice president of R&D for innovation transformation at Procter & Gamble, where he leads enterprise-wide work on how to create more meaningful innovation, faster. He has led

innovation programs across multiple business units and regions, including advancing product and package superiority through the integration of digital, data, and artificial intelligence capabilities. He is also a Harvard Business School, HBS AI Institute Industry Fellow.

Karim R. Lakhani is the Dorothy & Michael Hintze Professor of Business Administration at Harvard Business School, specializing in technology management, open innovation, and AI strategy and transformation. He is the founding chair of Harvard's Digital Data Design Institute and the Laboratory for Innovation Science. His work includes pioneering field experiments with organizations like NASA, Harvard Medical School, the Broad Institute, and Procter & Gamble. He holds a PhD in management from MIT.