Working Paper 25-043

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

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

Harvard Business School

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

Fabrizio Dell'Acqua
Harvard Business School and Digital Data Design Institute at Harvard

Charles Ayoubi
ESSEC Business School

Hila Lifshitz
Warwick Business School and Digital Data Design Institute at Harvard

Raffaella Sadun
Harvard Business School and Digital Data Design Institute at Harvard

Ethan Mollick
The Wharton School, University of Pennsylvania

Lilach Mollick
The Wharton School, University of Pennsylvania

Yi Han
Procter & Gamble

Jeff Goldman
Procter & Gamble

Hari Nair
Procter & Gamble

Stew Taub
Procter & Gamble

Karim R. Lakhani
Harvard Business School and Digital Data Design Institute at Harvard

Working Paper 25-043

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

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

Fabrizio Dell'Acqua[1,2], Charles Ayoubi[2,3], Hila Lifshitz[2,4], Raffaella Sadun[1,2], Ethan Mollick[5], Lilach Mollick[5], Yi Han[6], Jeff Goldman[6], Hari Nair[6], Stew Taub[6], and Karim R. Lakhani[1,2]

[1]Harvard Business School
[2]Digital Data Design Institute at Harvard
[3]ESSEC Business School
[4]Warwick Business School, Artificial Intelligence Innovation Network
[5]The Wharton School, University of Pennsylvania
[6]Procter & Gamble

March 28, 2025

Working Paper

*We thank Ramona Pop for her critical help managing the experiment. We thank Andrea Dorbu, Bandy Chin, Corey Gelb-Bicknell, Hadi Abbas, Michael Menietti, Sarah Stegall-Rodriguez and Vishnu Kulkarni for very helpful support and research assistance, and Brent Hecht for thoughtful comments. Funding for this research was provided in part by Harvard Business School. Procter & Gamble provided financial support to the Digital Data Design Institute (D^3) 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. We used Claude and ChatGPT for light copyediting. All errors are our own.

Abstract

We examine how artificial intelligence transforms the core pillars of collaboration—performance, expertise sharing, and social engagement—through a pre-registered field experiment with 776 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 reveal that AI significantly enhances performance: individuals with AI matched the performance of teams without AI, demonstrating that AI can effectively replicate certain benefits of human collaboration. Moreover, AI breaks down functional silos. Without AI, R&D professionals tended to suggest more technical solutions, while Commercial professionals leaned towards commercially-oriented proposals. Professionals using AI produced balanced solutions, regardless of their professional background. Finally, 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. Our results suggest that AI adoption at scale in knowledge work reshapes not only performance but also how expertise and social connectivity manifest within teams, compelling organizations to rethink the very structure of collaborative work.

Keywords: Artificial intelligence, Teamwork, Human-machine interaction, Productivity, Skills, Innovation, Field experiment.

1 Introduction

Teamwork is the cornerstone of modern organizations. Whether designing a new product, solving strategic challenges, or orchestrating large-scale innovation, human collaboration has traditionally been central to achieving higher-quality results than individuals working alone. There are three fundamental pillars upon which the justification for teamwork relies. The first is performance: teamwork is more effective than individual work and allows for more complex problems to be tackled (Ancona and Caldwell, 1992; Lindbeck and Snower, 2000; Wuchty et al., 2007; 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 in an effective way (Kogut and Zander, 1992; Argote, 1999; Nickerson and Zenger, 2004). Finally, human sociality: people enjoy connecting with other people, which increases their motivation to work (Deutsch, 1949; Kozlowski and Bell, 2013; Johnson and Johnson, 2005). Despite significant research on how teamwork and collaborations function, we know remarkably little about how these core pillars hold up when an emerging technology enters the equation: artificial intelligence (AI). The integration of AI into knowledge work poses a foundational challenge: while AI, particularly Generative AI (GenAI), has demonstrated the capacity to enhance individual creativity, productivity, and decision-making (Noy and Zhang, 2023; Dell'Acqua et al., 2023b; Brynjolfsson et al., 2025; Peng et al., 2023), 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 performance. But a unique aspect of Large Language Models, 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 the role of humans in teamwork? 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.

To address these questions, we designed a large-scale 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 certain specialized knowledge and skills? 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 unique field experiment and organizational upskilling program involving 776 experienced professionals at Procter & Gamble (P&G), a global consumer packaged goods company. Participants engaged in their company's standardized new product development process, randomly assigned to one of four conditions, in a 2x2 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. 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.

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 at scale, potentially with less time invested 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. While 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 replicates many of the benefits of human collaboration, acting as a "cybernetic teammate."[1] 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. Digging deeper, 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, effectively replicating the knowledge integration typically achieved through human collaboration. This has the potential to diminish functional boundaries, democratizing expertise within teams and organizations.

[1]The term draws from Norbert Wiener's 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 (Wiener, 1948, 1950).

Moreover, professionals reported more positive emotions and fewer negative emotions when engaging with AI compared to working alone, matching the emotional benefits traditionally associated with human teamwork. This pattern notably differs from previous findings about technology's typically negative impact on workplace social dynamics.

Overall, our findings indicate that adopting AI in knowledge work involves more than simply adding another tool. By enhancing performance, bridging functional expertise, and reshaping collaboration patterns, GenAI prompts a rethinking of how organizations structure teams and individual roles. As firms integrate AI technologies more widely, they must weigh not only operational efficiencies but also emotional and social implications for workers. Our study lays a foundation for understanding these shifts and offers insights that can guide the design of AI-enhanced work environments—where AI itself acts as a genuine teammate.

2 Related Literature

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). While 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 fundamentally 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, while 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. In this sense, teamwork serves as on-the-ground conduits of knowledge exchange, bridging cognitive gaps that would otherwise constrain performance.

Additionally, 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 (Garud, 1997; Kogut and Zander, 1992), whereas domain expertise focuses on the norms and application contexts that are unique to each sector. Both types of expertise can be crucial for surfacing and implementing innovative solutions effectively (Ayoubi et al., 2023).

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; Kozlowski and Bell, 2013; Johnson and Johnson, 2005). 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.

Against this backdrop of increasingly team-based knowledge work, GenAI has emerged as a transformative technology (Noy and Zhang, 2023; Dell'Acqua et al., 2023b; Brynjolfsson et al., 2025; Peng et al., 2023; Boussioux et al., 2025; Girotra et al., 2023; Doshi and Hauser, 2024;

Eloundou et al., 2024).[2] 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.

Generative AI 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 (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).

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 transform the division of labor by expanding the potential performance on certain tasks beyond what humans or AI could achieve on their own (Choudhary et al., 2023; Hoffmann et al., 2024). Finally, AI may also facilitate boundary-spanning across different knowledge domains (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). A recent laboratory study highlights these potential coordination pitfalls in human–AI partnerships (Dell'Acqua et al., 2023a). 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 (Trist and Bamforth, 1951; Henrich et al.,

[2]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); McElheran et al. (2024).

2001; Dell'Acqua et al., 2023a; Toner-Rodgers, 2024). Yet, recent meta-analytic evidence also 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 (Li et al., 2023, 2024). As a result, understanding not only whether AI can bolster performance but also how it shapes team expertise sharing and social interactions becomes a pressing topic for scholars and practitioners alike.

3 Experimental Design

3.1 Empirical Setting

Between May and July 2024, we conducted a large-scale field experiment at Procter & Gamble (P&G) to evaluate how GenAI influences cross-functional new product development.[3] P&G—renowned for its global footprint, structured R&D processes, and highly skilled workforce—provides an ideal environment 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 vast operational scope, offers a unique lens 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 fundamental unit where innovation happens in the company—it's 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

[3]This project (IRB24-0202) received IRB approval. The study was pre-registered at AEA RCT Registry (AEARCTR-0013603), detailing our experimental conditions, outcome variables, and analytical approaches. This plan will become publicly available upon article acceptance or after the registry's embargo period.

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

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.[4] We thus mimicked this structure in our experimental design.

The experiment was conducted as a one-day virtual product development workshop, involving 811 participants from P&G's Commercial and R&D functions.[5] Our analyses focus on 776 of these participants who were randomly assigned across four conditions.[6] Specifically, the four conditions were: (1) Control: Individual without AI, (2) Treatment 1 (T1): Team (R&D + Commercial) without AI, (3) Treatment 2 (T2): Individual + AI, and (4) Treatment 3 (T3): Team (R&D + Commercial) + 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).[7]

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

[5] The detailed description of the tasks given to participants can be found in Appendix.

[6] 35 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). Results are consistent when we include these non-randomized participants.

[7] While the randomization clusters included a geographical component, it was primarily to accommodate timezone

Randomization was stratified by business unit and geography to ensure balanced representation across all groups. Table 1 provides an overview of the participants, indicating a balanced distribution of key functions within P&G. Figure 1 illustrates our 2x2 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.[8] 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.[9] 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, 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. The GenAI tool used in the experiment was built on GPT-4 and accessed through Microsoft Azure.[10] 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 CPG-related tasks. One of the authors led this training and provided a PDF with recommended prompts. This standardized approach ensured a uniform baseline of familiarity with the GenAI interface for all AI-enabled participants. In addition to our primary measures

differences and ensure that team members could collaborate in real-time.

[8]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.

[9]The random assignment of leadership role between R&D and Commercial professionals had no statistically significant impact on any of our team outcomes.

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

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. Pre-survey data was collected to gather individual information about participants. During the product development workshop, all GenAI prompts and responses were recorded, and team interactions were transcribed. Post-survey data was also collected, and followup 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 managers, 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.

After completing their initial task, participants in the control groups (both individual and team) underwent the same GenAI training as the treated groups. They then repeated the task using the newly acquired AI skills, allowing for a within-participant comparison of performance before and after the training. This additional step not only provides insights into the learning curve associated with GenAI tools and their potential for rapid integration into existing work processes but also constitutes a cross-over experiment design for the control groups. It's important to note, however, that all the primary results presented in this study are based on the between subject comparisons, focusing on the initial performance across all conditions before any crossover occurred.

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 3.2 below.

Controls_i includes list of pre-experimental 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. Throughout our analysis, we use robust standard errors to account for potential heteroskedasticity.

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 proposed 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. Novelty measures the degree of innovation and originality in the proposed solutions on a scale from 1 to 10, while 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.[11]

These innovation outcomes are grounded in the literature (e.g., Lane (2023)) and 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 several other performance measures. Time Spent captures the number of seconds participants spent working on their task. In our analyses, we use the natural logarithm of Time Spent, as this transformation better accounts for the right-skewed nature of time measurements, though our results are consistent when using raw time values. 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 proposed 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% (1) or not (0). 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

[11] As a robustness check, we replicated all analyses using AI-generated evaluations of the solutions. Results remain consistent, as shown in Appendix.

might interact with AI or team structures. Additionally, we measured the degree of Technicality of a solution. A 1–7 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, while negative emotions aggregate feelings of anxiety, frustration, and distress. Both measures are calculated as the difference between post-task and pre-task responses, with each component measured on a scale from 1 to 7, and both measures are standardized based on the control group mean and standard deviation.

5 Results

5.1 Performance

Figure 2 provides crucial 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 more substantial: individuals with AI demonstrate a 0.37 standard deviation increase ($p < 0.01$), while teams with AI show a 0.39 standard deviation 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 large productivity improvements. Figure 4 illustrates the average time saved on tasks across different groups, using individuals without AI as the baseline. Teams and individuals without AI spent similar amounts of time on tasks. However, the introduction of AI substantially reduced time spent working on the solution: individuals with AI spent 16.4% less time than the control group, while teams with AI spent 12.7% less time. Table 4 further corroborates these findings. Additionally, Figure A1 shows the impact of AI was substantial. While teams without AI produced solutions only marginally longer than individual controls, the introduction of AI led to substantially 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 5 and the corresponding Table 5. These figures split our sample between employees for whom product development is a core job task (left panel - core-job) and employees that are less familiar with new product development (right panel - non-core-job), comparing their performance across our experimental conditions.[12]

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 democratize expertise within organizations, extending prior work on individual knowledge workers (e.g., (Brynjolfsson et al., 2025; Dell'Acqua et al., 2023b)). AI allows less experienced employees to achieve performance levels that previously required either direct collaboration or supervision by colleagues with more task-related experience.

[12]Teams where only one employee has as their core job to work on new product development are classified as core-job teams.

Our next findings focus on changes in the collaboration of teams. Figure 6 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 more commercial ideas, as indicated by their distribution towards higher values on the x-axis. In contrast, technical participants (yellow) tend to suggest less commercially-oriented ideas, clustering towards lower 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 6 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 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 7 and 8 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 6 and 7 confirm these results. Specifically, individuals with AI showed a 0.457 standard deviation increase in positive emotions ($p < 0.01$) compared to the control group, while 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 to those working in teams, reflecting the traditional psychological benefits of human collaboration. However, individuals using AI report positive emotional responses that match or 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 individual work settings.

These emotional responses correlate with participants' evolving expectations about AI use. As shown in Tables 8 and 9, participants who reported larger increases in their expected future use of AI also reported more positive and fewer negative emotions during the task. While this correlation cannot definitely establish causality, it suggests an interesting relationship between positive experiences with AI and anticipated future engagement with the technology.

6 Additional Analyses

6.1 Exceptional Performance Measures

While our primary analyses center on average solution quality, many organizations place disproportionate emphasis on exceptional outcomes—the very best ideas that may generate outsized returns if implemented. 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). Understanding how different work configurations affect the likelihood of generating these exceptional solutions is therefore crucial for organizations seeking to optimize their innovation processes.

To explore whether AI can facilitate these standout solutions, we developed additional metrics capturing top-tier performance. We created a binary measure called Top 10% Solutions, which equals 1 if a solution's quality score (on a 1–10 scale) ranked in the highest decile across all submissions in the sample, and 0 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 using AI were more likely to generate solutions ranking in the top 10% of all submissions. Specifically, as quantified in Table 10, teams with AI were 9.2 percentage points more likely to produce solutions in the top decile compared to the control mean of 5.8%, that corresponds to around 3 times more chances of being in the top decile of solutions. While 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, doesn't just improve average performance but substantially increases the likelihood of producing the kind of breakthrough solutions that drive organizational success.

6.2 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, while 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 to the control group ($p < 0.05$), suggesting a disconnect between actual and perceived performance.

6.3 Human Team Collaboration

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

6.4 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 towards 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.[13] Interestingly, the distribution also shows a non-trivial 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.

Considering more broadly the variety of ideas being produced, Figure 13 exhibits the semantic similarity of solutions across different conditions. While human-only solutions (both individual and pair) show relatively dispersed distributions, AI-aided solutions demonstrate notably higher semantic similarity. This increased consistency in AI-aided solutions aligns with existing literature on the standardizing effect of large language models. However, in order to better interpret the similarity increase, we directly prompted GPT-4o to solve the same problems iteratively and checked whether AI-enabled solutions were especially similar to what AI alone generated.[14] This "AI Only" shows much tighter clustering, suggesting that human participants are not simply transcribing naive AI outputs. This finding becomes particularly interesting when considered alongside our retention analysis: despite the high retention rates of AI-generated content in final solutions, the semantic fingerprint of AI-aided solutions remains closer to human-only solutions than to pure AI outputs, indicating that humans meaningfully shape and contextualize AI suggestions rather than merely adopting them wholesale.

[13] Among participants who retained at least some AI-generated content, the average number of prompts was 18.7. Notably, participants whose solutions showed 100% AI-generated content averaged 23.9 prompts, suggesting extensive iterative interaction with the tool rather than simple copy-and-paste behavior.

[14] We simply prompted the GPT-4o interface with the instructions of the problem with no additional iterations.

7 Discussion and Conclusion

Our study reveals fundamental insights about the transformative potential of GenAI in workplace team collaboration, with implications for both theory and practice. Our findings demonstrate that AI integration is not merely augmenting existing work processes but may have the potential to reshape 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 improvement) compared to individuals working alone, reflecting the traditional benefits of cross-functional collaboration. However, the introduction of AI dramatically reshapes this performance landscape. Individuals working with AI showed a substantial 0.37 standard deviation performance increase over the baseline of working alone without AI. 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 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 from fundamentally transforming human-to-human collaboration.

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.

We should also highlight two 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, while collaborations involving team members with similar expertise, or in larger, more intricate team structures, may exhibit different patterns of AI adoption and effectiveness.

Perhaps our most striking finding concerns AI's role in transforming professional expertise boundaries. Traditional organizational theory has long emphasized the importance of specialized knowledge and clear functional boundaries. Our results suggest AI fundamentally disrupts this paradigm. Without AI, we observed clear professional silos - Commercial specialists proposed predominantly commercial solutions while R&D professionals favored technical approaches. When teams worked without AI, they produced more balanced solutions through cross-functional collaboration. Remarkably, 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.

The emotional implications of AI integration 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, GenAI's interactive features appear to create remarkably positive experiences for workers, aligning with emerging evidence on the beneficial psychological effects of conversational AI (Trist and Bamforth, 1951; Dell'Acqua et al., 2023a; Li et al., 2023; De Freitas et al., 2024). These findings suggest that successful AI integration should focus on helping workers better recognize and internalize their improved performance capabilities.

These results indicate that AI is no longer merely a passive tool but rather functions as a "cybernetic teammate." By interfacing dynamically with human problem-solvers — providing real-time feedback, bridging cross-functional expertise, and influencing self-reported emotional states — GenAI shows its capacity to occupy roles we typically associate with human collaborators. In this sense, AI not only enhances individual cognitive work but also replicates key collective functions, such as ideation and iterative refinement, helping teams address complex challenges more holistically. While AI cannot fully replicate the richness of human social and emotional interaction, its ability to contribute as a genuine collaborator suggests a marked shift in how knowledge work can be structured and carried out.[15] Our findings also speak to a growing body of literature that conceptualizes AI not merely as a tool or a medium, but rather as an active "counterpart" within broader socio-technical systems. Drawing on distributed cognition

[15]See Leonardi and Neeley (2022) and Farrell et al. (2025) for related discussions.

(Hutchins, 1991, 1995) and Actor–Network Theory (Callon, 1984; Latour, 1987, 2007), recent organizational work highlights the importance of examining AI's development, implementation, and use alongside a wide array of human actors and organizational infrastructures (Anthony et al., 2023). Our study supports and extends these arguments by demonstrating that GenAI can shape expertise sharing, team dynamics, and social engagement in ways that exceed the traditional boundaries of automation. In other words, AI's role transcends that of a mere tool or facilitator, entering the relational fabric of collaboration itself. By treating AI as an active counterpart, and in fact as a proper teammate, we gain deeper insight into how GenAI mediates, and is mediated by, the collective processes that form the backbone of modern teamwork.

These findings have significant organizational implications. First, organizations may need to fundamentally rethink 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, while others may focus on the complementarity.[16] The increased speed and comprehensiveness of AI-enabled work—evidenced by significantly longer solutions produced in less time—suggests 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 is also value in training workers to think more broadly across functional boundaries.

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, what features of AI systems specifically support effective knowledge integration across professional boundaries? Third, how do organizations effectively capture and disseminate best practices for AI-enabled work? Finally, how does AI integration affect the development of domain expertise over time? Does AI-enabled boundary spanning lead to genuine expertise development, or does it primarily facilitate access to existing knowledge?

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

Our research demonstrates that AI adoption necessitates rethinking fundamental assumptions about team structures and organizational design. By showing that AI can elevate individual performance to levels comparable to traditional teams while simultaneously breaking down 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 complex interactions between human and artificial capabilities that merit further investigation. As organizations continue to integrate AI technologies, understanding these dynamics will be crucial 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.

These findings challenge the notion of AI as merely an advanced search engine or convenient text generator, instead highlighting its role as an active participant in collaborative networks. By contributing to decision-making, creativity, and even emotional responses, AI is reshaping the conditions under which teams form and function. While questions remain about how AI will influence long-term skill development and trust, our evidence underscores a pivotal shift in knowledge work—one that calls for new ways of understanding the evolving interplay between human and machine contribution, and a new science of cybernetic teams

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Prediction machines: the simple economics of artificial intelligence*, Harvard Business Press, 2018.

Alchian, Armen A and Harold Demsetz, "Production, information costs, and economic organization," *The American Economic Review*, 1972, *62* (5), 777–795.

Ancona, Deborah G. and David F. Caldwell, "Bridging the boundary: External activity and performance in organizational teams," *Administrative Science Quarterly*, 1992, *37* (4), 634–665.

Anthony, C., B. A. Bechky, and A. L. Fayard, ""Collaborating" with AI: Taking a system view to explore the future of work," *Organization Science*, 2023.

Argote, Linda, "Organizational learning: creating," *Retaining and Transferring*, 1999, *25*, 45–58.

_ , Sunkee Lee, and Jisoo Park, "Organizational learning processes and outcomes: Major findings and future research directions," *Management Science*, 2021, *67* (9), 5399–5429.

Ayoubi, Charles, Jacqueline N Lane, Zoe Szajnfarber, and Karim R Lakhani, "The Dual Effect of Intellectual Similarity: The Interplay of Critique and Favoritism in the Evaluation of Technological Innovations," *Working paper*, 2023.

_ , Michele Pezzoni, and Fabiana Visentin, "At the origins of learning: Absorbing knowledge flows from within the team," *Journal of Economic Behavior & Organization*, 2017, *134*, 374–387.

Balasubramanian, Natarajan, Yang Ye, and Mingtao Xu, "Substituting human decision-making with machine learning: Implications for organizational learning," *Academy of Management Review*, 2022, *47* (3), 448–465.

Beane, Matthew, "Shadow learning: Building robotic surgical skill when approved means fail," *Administrative Science Quarterly*, 2019, *64* (1), 87–123.

Boudreau, Kevin J, Nicola Lacetera, and Karim R Lakhani, "Incentives and problem uncertainty in innovation contests: An empirical analysis," *Management Science*, 2011, *57* (5), 843–863.

Boussioux, Leonard, Jacqueline N Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R. Lakhani, "The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing," *Organization Science*, 2025, *35* (5).

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson, "Artificial intelligence and the modern productivity paradox," in "The economics of artificial intelligence: An agenda," Vol. 23, University of Chicago Press, 2019, pp. 23–57.

_ , Danielle Li, and Lindsey R. Raymond, "Generative AI at work," *The Quarterly Journal of Economics*, 2025.

_ , Tom Mitchell, and Daniel Rock, "What can machines learn and what does it mean for occupations and the economy?," in "AEA papers and proceedings," Vol. 108 American Economic Association 2018, pp. 43–47.

Callon, Michel, "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay," in John Law, ed., *Power, action and belief: A new sociology of knowledge?*, Routledge, 1984, pp. 196–223.

Cattani, Gino, Simone Ferriani, and Andrea Lanza, "Deconstructing the outsider puzzle: The legitimation journey of novelty," *Organization Science*, 2017, *28* (6), 965–992.

Choudhary, Vivek, Arianna Marchetti, Yash Raj Shrestha, and Phanish Puranam, "Human-AI ensembles: When can they work?," *Journal of Management*, 2023, p. 01492063231194968.

Cohen, Susan G and Diane E Bailey, "What makes teams work: Group effectiveness research from the shop floor to the executive suite," *Journal of Management*, 1997, *23* (3), 239–290.

Csaszar, Felipe A, "Organizational structure as a determinant of performance: Evidence from mutual funds," *Strategic Management Journal*, 2012, *33* (6), 611–632.

Dahan, Ely and Haim Mendelson, "An extreme-value model of concept testing," *Management Science*, 2001, *47* (1), 102–116.

Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski, "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination," *The Review of Economics and Statistics*, 2023.

_, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality," Working Paper 24-013, Harvard Business School Technology & Operations Mgt. Unit 2023.

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Deutsch, Morton, "A theory of co-operation and competition," *Human Relations*, 1949, *2* (2), 129–152.

DiBenigno, Julia and Katherine C Kellogg, "Beyond occupational differences: The importance of cross-cutting demographics and dyadic toolkits for collaboration in a US hospital," *Administrative Science Quarterly*, 2014, *59* (3), 375–408.

Doshi, Anil R. and Oliver P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science Advances*, 2024, *10* (28), eadn5290.

Dougherty, Deborah, "Interpretive barriers to successful product innovation in large firms," *Organization Science*, 1992, *3* (2), 179–202.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "Gpts are gpts: An early look at the labor market impact potential of large language models," *Science*, 2024, *384* (6702).

Farrell, Henry, Alison Gopnik, Cosma Shalizi, and James Evans, "Large AI models are cultural and social technologies," *Science*, 2025, *387* (6739), 1153–1156.

Freitas, Julian De, Ahmet Kaan Uguralp, Zeliha Uguralp, and Stefano Puntoni, "AI companions reduce loneliness," *SSRN Working Paper*, 2024.

Furman, Jason and Robert Seamans, "AI and the Economy," *Innovation policy and the economy*, 2019, *19* (1), 161–191.

Garud, Raghu, "Know-how, know-why, and know-what," *Advances in Strategic Management*, 1997, *14*, 81–101.

Girotra, K., L. Meincke, C. Terwiesch, and K. T. Ulrich, "Ideas are dimes a dozen: Large language models for idea generation in innovation," 2023. Available at SSRN: https://ssrn.com/abstract=4526071.

Girotra, Karan, Christian Terwiesch, and Karl T Ulrich, "Idea generation and the quality of the best idea," *Management Science*, 2010, *56* (4), 591–605.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath, "Cooperation, reciprocity and punishment in fifteen small-scale societies," *American Economic Review*, 2001, *91* (2), 73–78.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu, "Generative AI and the Nature of Work," Technical Report, CESifo Working Paper 2024.

Hutchins, Edwin, "Organizing work by adaptation," *Organization Science*, 1991, *2* (1), 14–39.

— , *Cognition in the wild*, Cambridge, MA: MIT Press, 1995.

Iansiti, Marco and Karim R. Lakhani, *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*, Harvard Business Press, 2020.

Jacobides, Michael G., Stefano Brusoni, and Francois Candelon, "The evolutionary dynamics of the artificial intelligence ecosystem," *Strategy Science*, 2021, *6* (4), 412–435.

Johnson, David W and Roger T Johnson, "New developments in social interdependence theory," *Genetic, Social, and General Psychology Monographs*, 2005, *131* (4), 285–358.

Jones, Benjamin F, "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?," *The Review of Economic Studies*, 2009, *76* (1), 283–317.

Kacperczyk, Aleksandra and Peter Younkin, "The paradox of breadth: The tension between experience and legitimacy in the transition to entrepreneurship," *Administrative Science Quarterly*, 2017, *62* (4), 731–764.

Kellogg, Katherine C, Wanda J Orlikowski, and JoAnne Yates, "Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations," *Organization Science*, 2006, *17* (1), 22–44.

Kogut, Bruce and Udo Zander, "Knowledge of the firm, combinative capabilities, and the replication of technology," *Organization science*, 1992, *3* (3), 383–397.

Kozlowski, Steve WJ and Bradford S Bell, *Work groups and teams in organizations: Review update.* 2013.

Lane, Jacqueline N, "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'," *Research Policy*, 2023, *52* (3), 104707.

Latour, Bruno, *Science in action: How to follow scientists and engineers through society*, Cambridge, MA: Harvard University Press, 1987.

—, *Reassembling the social: An introduction to actor-network-theory*, Oxford: Oxford University Press, 2007.

Lazer, David and Nancy Katz, "Building effective intra-organizational networks: The role of teams," *Working Paper*, 2003.

Leonardi, Paul and Tsedal Neeley, *The Digital Mindset: What It Really Takes to Thrive in the Age of Data, Algorithms, and AI*, Boston, MA: Harvard Business Review Press, 5 2022.

Levina, Natalia and Emmanuelle Vaast, "The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems," *MIS Quarterly*, 2005, *29* (2), 335–363.

Li, Han, Renwen Zhang, Yi-Chieh Lee, Robert E Kraut, and David C Mohr, "Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being," *NPJ Digital Medicine*, 2023, *6* (1), 236.

Li, Joanna Zun, Alina Herderich, and Amit Goldenberg, "Skill but not effort drive GPT overperformance over humans in cognitive reframing of negative scenarios," *Working Paper*, 2024.

Lindbeck, Assar and Dennis J. Snower, "Multitask learning and the reorganization of work: From tayloristic to holistic organization," *Journal of Labor Economics*, 2000, *18* (3), 353–376.

March, James G. and Herbert A. Simon, *Organizations*, New York: John Wiley & Sons, 1958.

McElheran, Kristina, J. Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI adoption in America: Who, what, and where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415.

Mollick, Ethan, *Co-Intelligence*, Random House UK, 2024.

Nelson, Richard and Sidney Winter, *An Evolutionary Theory of Economics Change* 1982.

Nickerson, Jack A and Todd R Zenger, "A knowledge-based theory of the firm—The problem-solving perspective," *Organization Science*, 2004, *15* (6), 617–632.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6684).

Page, Scott E., *The diversity bonus: How great teams pay off in the knowledge economy*, Princeton University Press, 2019.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer, "The impact of ai on developer productivity: Evidence from github copilot," *arXiv preprint arXiv:2302.06590*, 2023.

Puranam, Phanish, *The Microstructure of Organizations*, New York: Oxford University Press, 2018.

Raisch, Sebastian and Sebastian Krakowski, "Artificial intelligence and management: The automation–augmentation paradox," *Academy of Management Review*, 2021, *46* (1), 192–210.

Raj, Manav and Robert Seamans, "Primer on artificial intelligence and robotics," *Journal of Organization Design*, 2019, *8* (1), 1–14.

Souitaris, Vangelis, Bo Peng, Stefania Zerbinati, and Dean A Shepherd, "Specialists, generalists, or both? Founders' multidimensional breadth of experience and entrepreneurial ventures' fundraising at IPO," *Organization Science*, 2023, *34* (2), 557–588.

Toner-Rodgers, Aidan, "Artificial intelligence, scientific discovery, and product innovation," *arXiv preprint arXiv:2412.17866*, 2024.

Trist, Eric Lansdown and Ken W Bamforth, "Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system," *Human relations*, 1951, *4* (1), 3–38.

Weber, Roberto A. and Colin F. Camerer, "Cultural conflict and merger failure: An experimental approach," *Management Science*, 2003, *49* (4), 400–415.

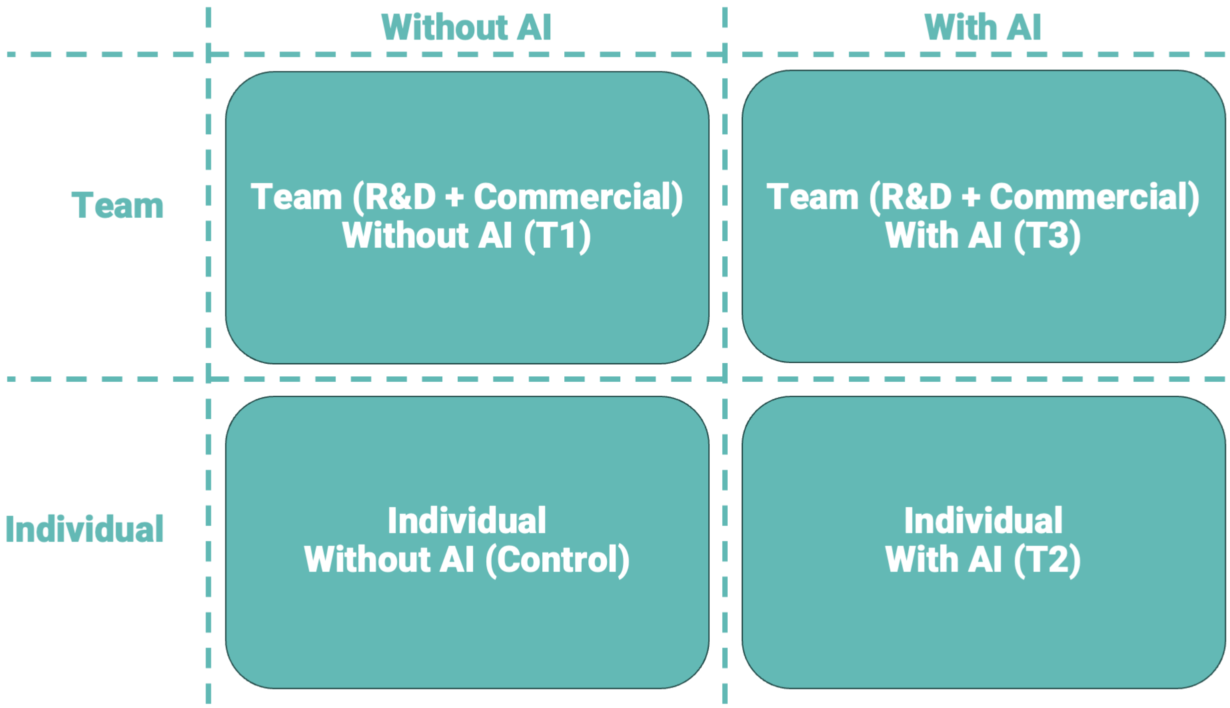
Weidmann, Ben and David J. Deming, "Team Players: How Social Skills Improve Group Performance," *National Bureau of Economic Research*, 2020, *w27071.*

Wiener, Norbert, *Cybernetics: Or Control and Communication in the Animal and the Machine*, Cambridge, MA: MIT Press, 1948.

__ , *The Human Use of Human Beings: Cybernetics and Society*, Boston: Houghton Mifflin, 1950. First Edition.

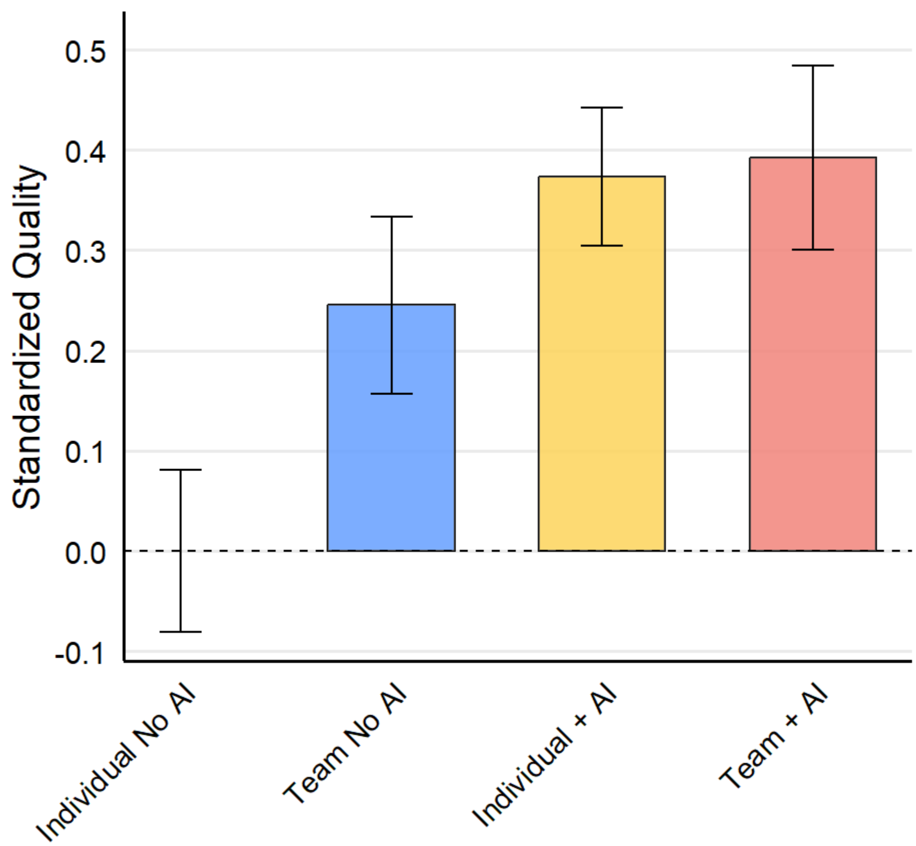
Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, 2007, *316* (5827), 1036–1039.

Figure 1: **Treatment Matrix**



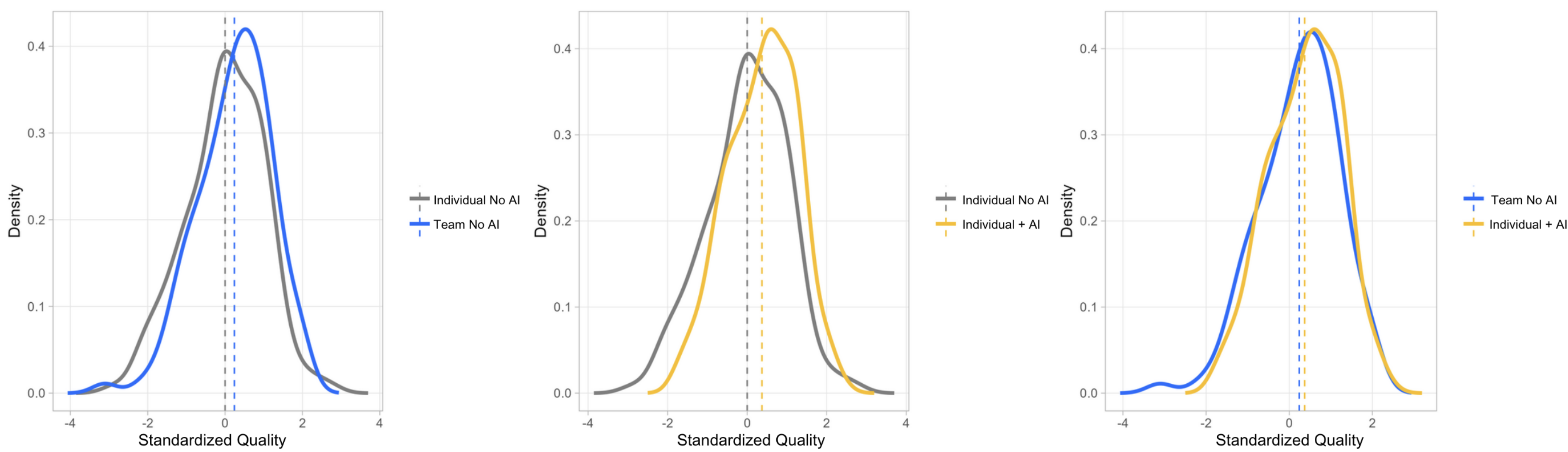
Notes: This figure displays the 2x2 experimental design showing four conditions: individuals and teams working either with or without AI assistance.

Figure 2: **Average Solution Quality**



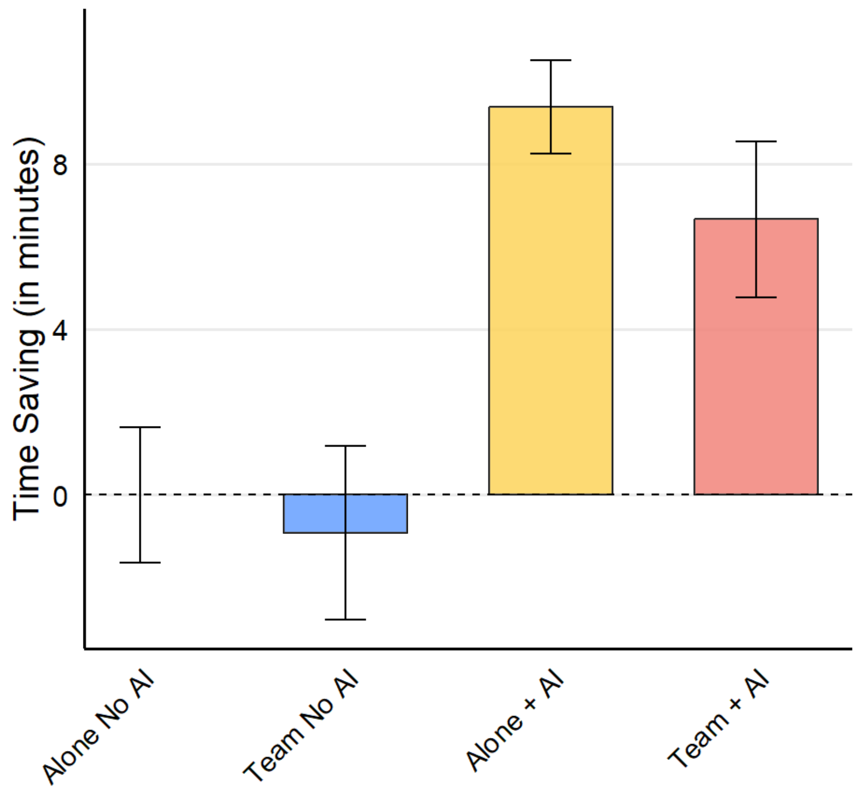
Notes: This figure displays the average quality scores for solutions across different groups, showing the relative performance of AI-treated versus non-AI-treated groups with standard errors.

Figure 3: **Pairwise Density Comparisons**



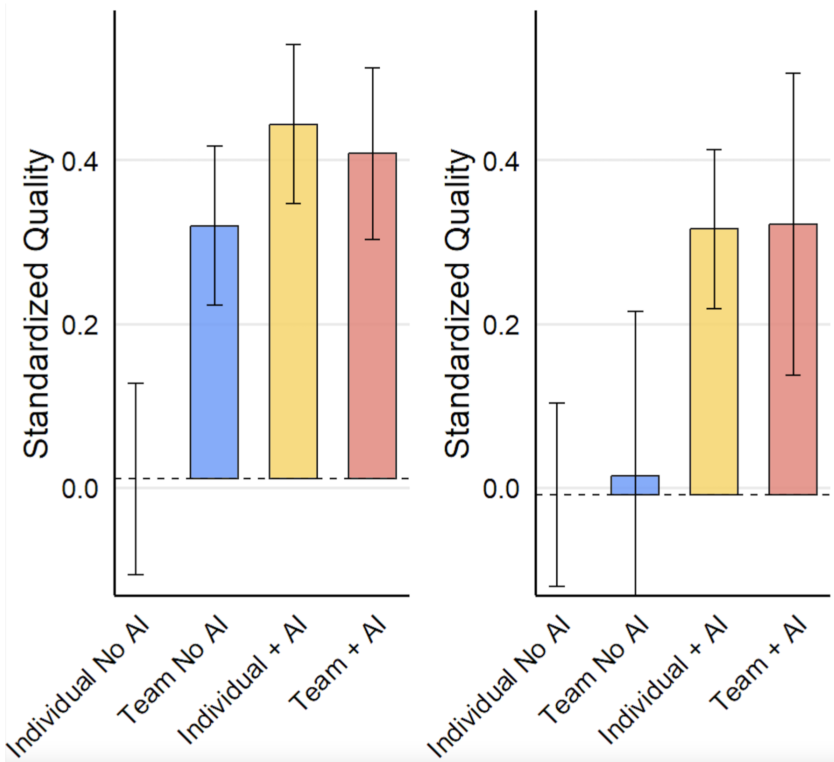
Notes: These figures illustrate the pairwise comparisons of solution quality distributions across different experimental conditions. The left panel compares solutions between individuals and teams working without AI assistance. The middle panel shows the quality distribution between individuals working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

Figure 4: **Time Saved**



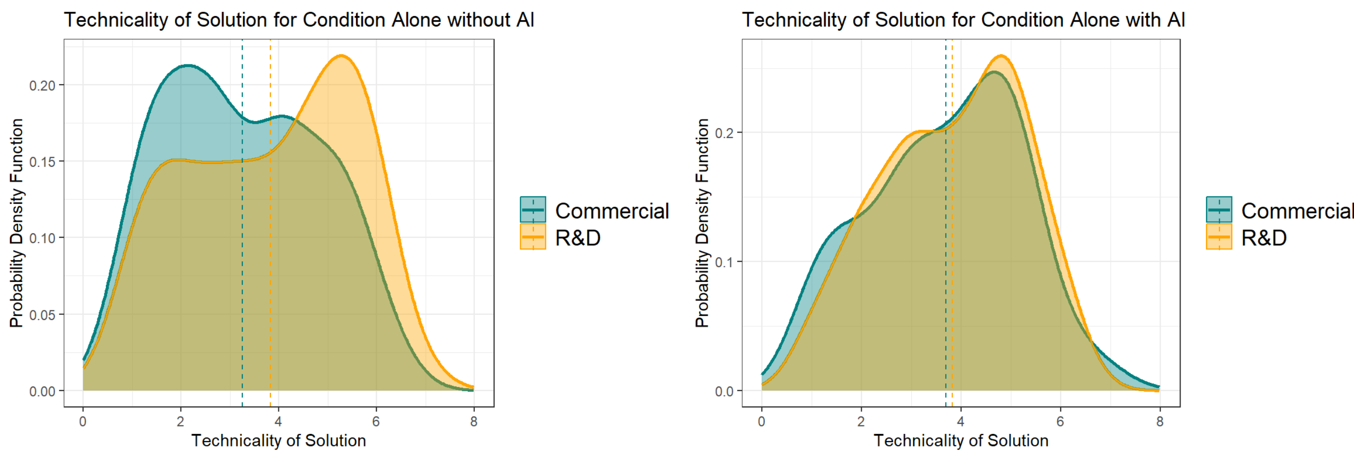
Notes: This figure shows the average time saved (in minutes) when preparing solutions by groups treated with AI versus those without AI with standard errors.

Figure 5: **Average Solution Quality: Core-jobs versus Not**



Notes: 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.

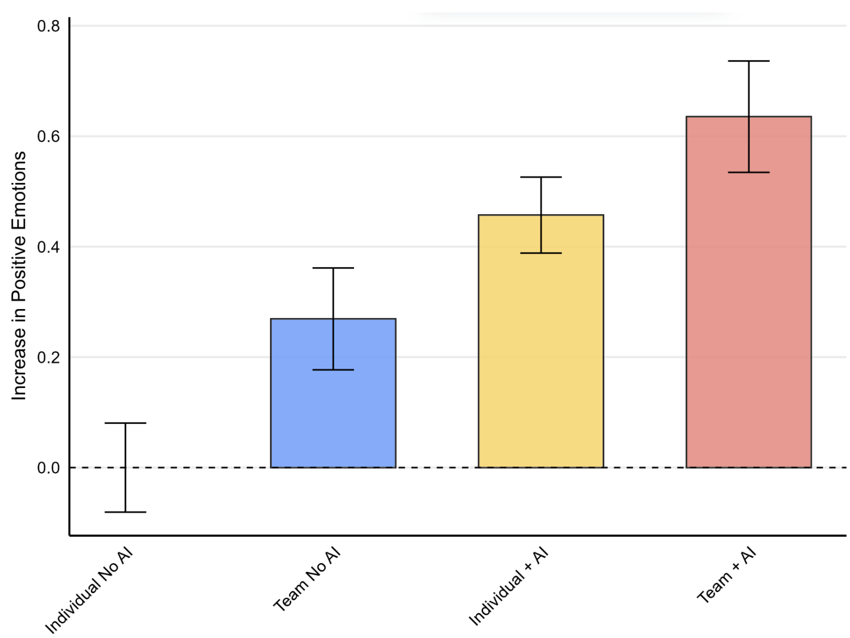
Figure 6: **Degree of Solution Technicality for Individuals**



(a) Individual - No AI (b) Individual - With AI

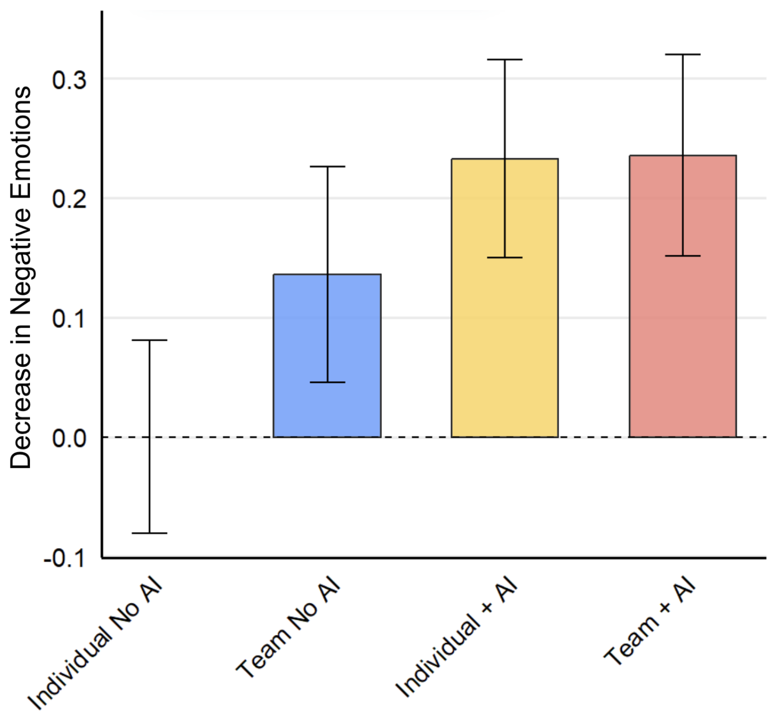
Notes: These figures illustrate the difference in idea generation between commercial and technical participants, with and without AI assistance. In both graphs, blue represents commercial participants and yellow represents technical participants. The x-axis indicates the commercial nature of ideas, with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

Figure 7: **Evolution of Positive Emotions during the Task**



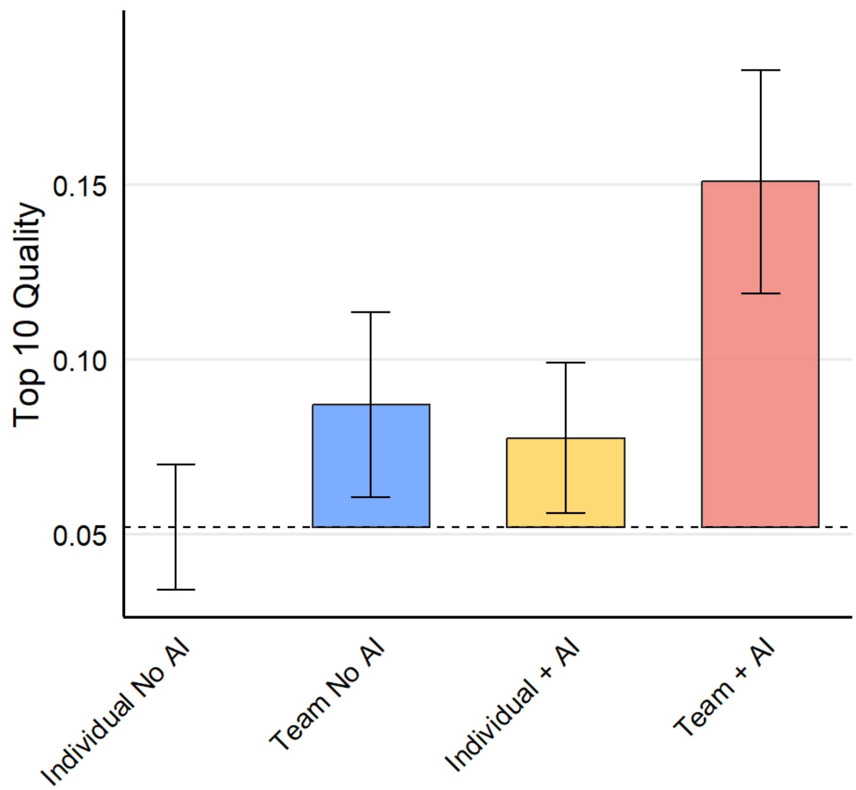
Notes: This figure presents the difference in self-reported positive emotions among participants before and after the task, comparing AI-treated and non-AI-treated groups to examine the emotional impact of AI on teamwork with standard errors. Positive emotions are answers to questions about enthusiasm,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

Figure 8: **Evolution of Negative Emotions during the Task**



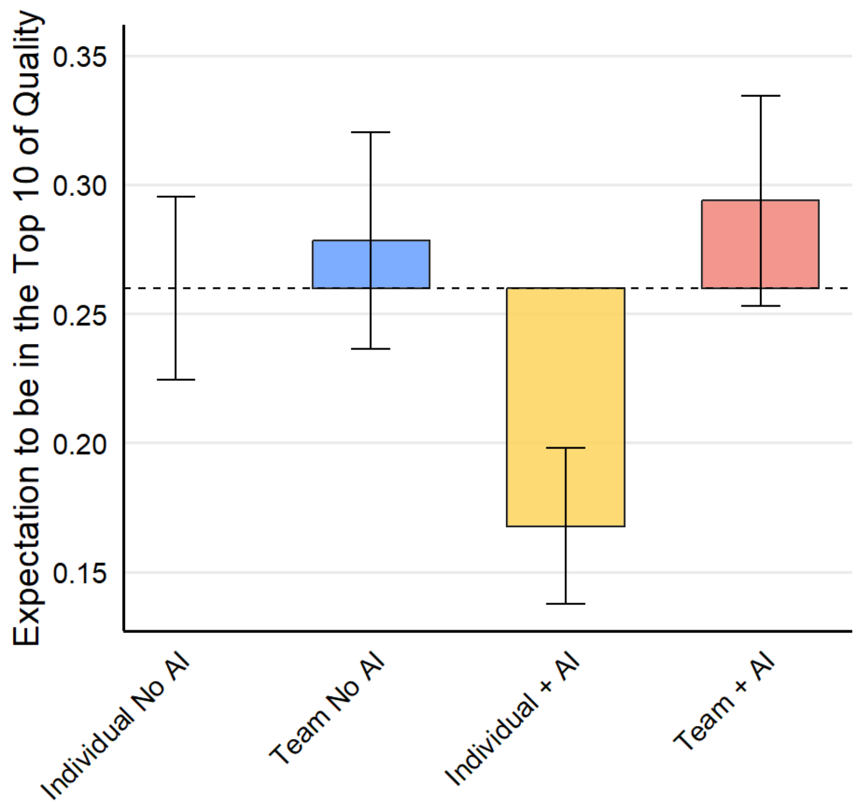
Notes: This figure presents the reduction in self-reported negative emotions among participants before and after the task, comparing AI-treated and non-AI-treated groups to examine the emotional impact of AI on teamwork with standard errors. Negative emotions are answers to questions about anxiety, frustration,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

Figure 9: **Top 10% Solutions**



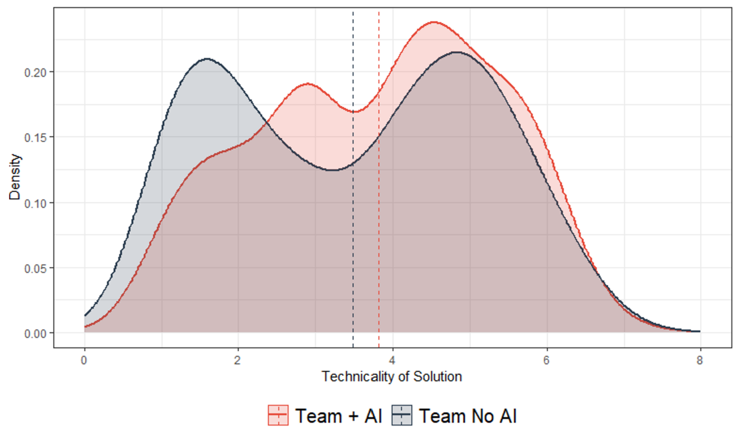
Notes: This figure displays the proportion of top 10% solution across different treatments with standard errors.

Figure 10: **Perceived Likelihood of Top 10 Placement by Treatment Group**



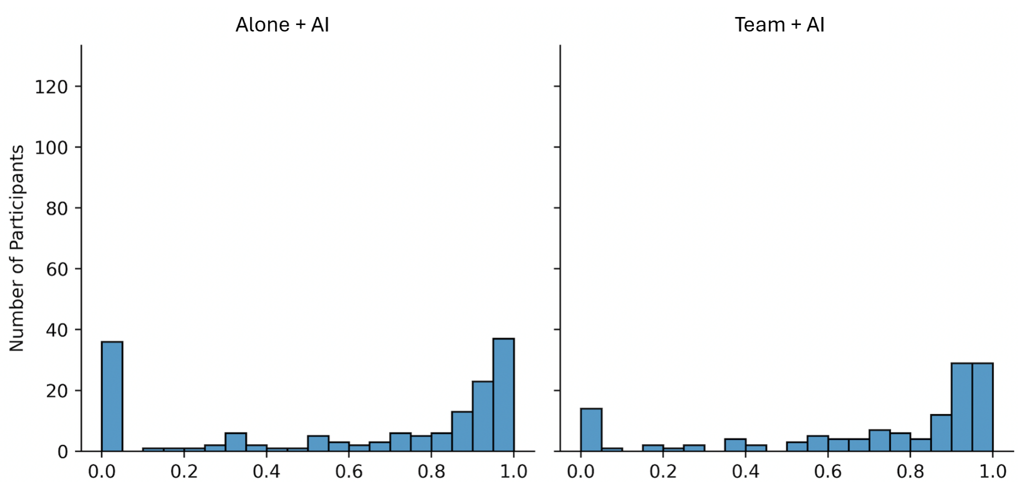
Notes: This table shows the percentage of participants in each treatment group who expected their solution to rank among the top 10. It reflects participants' confidence in their solutions across different conditions with standard errors.

Figure 11: **Degree of Solution Technicality for Teams**



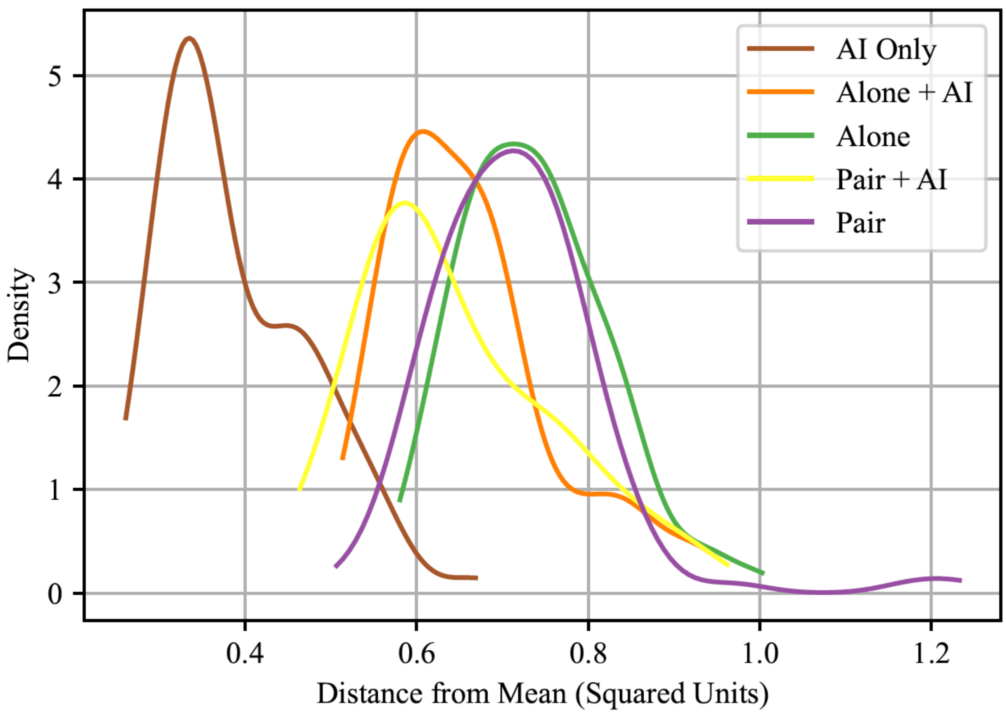
Notes: These figures illustrate the difference in idea generation for teams. Dark blue represents Team No AI and red represents Team + AI. The x-axis indicates the commercial nature of ideas, with higher values representing more technically-oriented suggestions.

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). Retainment 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.

Figure 13: **Similarity between Solutions**



Notes: This figure shows the kernel density distribution of semantic similarity across solution types. Distance from mean represents how semantically different solutions are from each other within each condition, with lower values indicating higher similarity. We measure semantic similarity using sentence embeddings and calculate the cosine distance between solutions.

Table 1: Summary Statistics

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 (1-5 Likert)	3.539 (0.951)	3.555 (1.027)	0.016
Individuals	154	155	
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 (1-5 Likert)	3.430 (1.003)	3.534 (1.021)	0.103
Team participants	230 (115 Teams)	252 (126 Teams)	

Note: Standard deviations in parentheses. + $p < 0.2$, $* \, p < 0.1$, $** \, p < 0.05$, $*** \, p < 0.01$

Table 2: Solution Quality (Standardized)

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

Note: P-values for the t-tests comparing "Team+AI" and "Team No AI" are reported. Fixed effects and controls as discussed in the text. $+\ p < 0.2$, $*\ p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

Table 3: Solution Length

	Length	Length	Length
Team No AI	30.456	56.746*	57.184+
	(27.419)	(30.865)	(38.673)
Individual + AI	504.507***	511.568***	503.833***
	(42.963)	(45.206)	(45.081)
Team + AI	543.745***	556.997***	551.578***
	(42.328)	(43.737)	(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

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+p<0.2$, $*p<0.1$, $**p<0.05$, $***p<0.01$

Table 4: Total Time for Task Completion (Log)

	Log Time	Log Time	Log Time
Team No AI	0.038	0.023	−0.015
	(0.072)	(0.072)	(0.080)
Individual + AI	-0.366^{***}	-0.374^{***}	-0.362^{***}
	(0.070)	(0.070)	(0.070)
Team + AI	-0.318^{***}	-0.324^{***}	-0.344^{***}
	(0.078)	(0.078)	(0.090)
Team+AI = Individual+AI	$p = 0.539$	$p = 0.519$	$p = 0.467$
Fixed Effects		X	X
Controls			X
Control Mean	7.333	7.548	7.666
Observations	550	550	550
Adjusted R^2	0.075	0.098	0.112

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+p<0.2$, $*p<0.1$, $**p<0.05$, $***p<0.01$

Table 5: Solution Quality by Familiarity with the Type of Task (Standardized)

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

+ $p < 0.2$, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses. Fixed effects and controls as discussed in the text. + $p < 0.2$, $*$ $p < 0.1$, **$p < 0.05$, $***$ $p < 0.01$

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

	Positive Emotions	Positive Emotions	Positive Emotions
Team No AI	0.269**	0.254**	0.257*
	(0.124)	(0.126)	(0.137)
Individual + AI	0.457***	0.475***	0.485***
	(0.107)	(0.106)	(0.106)
Team + AI	0.635***	0.635***	0.666***
	(0.131)	(0.129)	(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

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+\ p < 0.2$, $*\ p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

Table 7: Evolution of Self-Reported Negative Emotions Before and After the Task (Standardized)

	Negative Emotions	Negative Emotions	Negative Emotions
Team No AI	−0.136	−0.094	−0.006
	(0.124)	(0.121)	(0.141)
Individual + AI	−0.233**	−0.247**	−0.263**
	(0.117)	(0.116)	(0.117)
Team + AI	−0.235**	−0.221*	−0.157
	(0.118)	(0.116)	(0.138)
Fixed Effects		X	X
Controls			X
Control Mean	0.000	0.166	0.068
	(0.082)	(0.166)	(0.252)
Observations	530	530	530
Adjusted R^2	0.005	0.022	0.031

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+\ p < 0.2$, $*\ p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

Table 8: 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 E.	Positive E.	Positive E.	Positive E.	Positive E.	Positive E.
Diff. in Expected Use of ChatPG	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

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+p<0.2$, $*p<0.1$, $**p<0.05$, $***p<0.01$

Table 9: 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 E.	Negative E.	Negative E.	Negative E.	Negative E.	Negative E.
Diff. in Expected Use of ChatPG	-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

Note: Standard errors in parentheses. Fixed effects and controls as discussed in the text. $+p<0.2$, $*p<0.1$, $**p<0.05$, $***p<0.01$

Table 10: Probability of Being Rated Top 10% of Quality Scores

	Top Quality	Top Quality	Top Quality
Team No AI	0.037	0.045^{+}	0.054^{+}
	(0.033)	(0.034)	(0.041)
Individual + AI	0.019	0.029	0.030
	(0.029)	(0.029)	(0.029)
Team + AI	0.092^{**}	0.098^{**}	0.112^{**}
	(0.037)	(0.038)	(0.045)
Team+AI = Team No AI	p = 0.190	p = 0.207	p = 0.175
Team+AI = Individual+AI	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

Note: P-values for the t-tests comparing "Team+AI" with "Team No AI" and "Individual+AI" are reported. Fixed effects and controls as discussed in the text. $+p<0.2$, $*p<0.1$, $**p<0.05$, $***p<0.01$

Appendix

A Problem Statements

We report below the problem statements presented to participants during the hackathon. These statements reflected real business challenges that the respective business units were actively working on at the time of the experiment. Each statement was accompanied by relevant market data and additional contextual information provided by the business units. All statements represented significant innovation opportunities identified by senior management. For confidentiality, we have removed specific brand names and company references, indicated by [brand] or [company].

1. **Business Unit 1 Problem Statement:**
   "How to help consumers transition from product form X to Y [specific product examples removed]?"

2. **Business Unit 2 Problem Statement:**
   "How to motivate consumers who have never tried product form X to try it as part of their regimen"

3. **Business Unit 3 Problem Statement:**
   "How do we make the current portfolio of Brand X form/regimen offerings in the category simple to understand and choose to shop as a 'one size fits all solution', versus competitors who offer only a single offering?[company and competitor examples removed])?"

4. **Business Unit 4 Problem Statement:**
   "What are ways we can affect the consumer dosing habits of product X to help them achieve better health?"

B Solution Evaluation Process

This section details the evaluation process used to assess the quality and characteristics of solutions generated during the experiment.

B.1 Evaluator Selection and Composition

The evaluation of solutions was conducted by a panel of 22 expert evaluators who collectively performed 1,595 evaluations across 550 unique solutions, resulting in an average of around three evaluations per solution. All evaluators were experienced professionals with backgrounds in business and technology - MBA and Engineering students, or recent graduates, at a top business or engineering school, ensuring a comprehensive assessment of both technical and commercial aspects of the proposed solutions.

B.2 Evaluation Process

Each evaluator was assigned approximately 70 solutions to review. For each solution, evaluators assessed the solutions, comprising of five key components:

1. Idea Name
2. Recommended Solution
3. Rationale Details
4. Critical Work Required
5. Support or Resources Needed for Implementation

B.3 Evaluation Metrics

Evaluators assessed each solution on five primary dimensions using a 1-10 scale:

- **Overall Quality**: A comprehensive assessment of the solution's merit
- **Novelty**: The originality and uniqueness of the approach
- **Impact**: The effectiveness in addressing the problem and creating value
- **Business Potential**: The potential for significant business benefit and value creation
- **Feasibility**: The practicality and achievability of the proposed approach

Additionally, evaluators assessed the technical versus commercial orientation of each solution on a separate 1-7 Likert scale.

B.4 Quality Control and Evaluation Reliability

To maintain evaluation integrity, evaluators agreed to strict confidentiality requirements.

To ensure evaluation reliability, solutions received multiple independent assessments. Final scores for each solution were calculated by averaging all individual evaluations.

To assess evaluation consistency, we measured inter-rater reliability using multiple metrics. Our analysis revealed an ICC2 of 0.452, Kendall's Tau of 0.153, and Pearson's r of 0.198. These values align with established reliability standards in innovation assessment, where Seeber et al. (2024) report ICC values of 0.11-0.55 for grant evaluations. The variance distribution (total: 3.93; solution: 1.77; evaluator: 0.51; error: 1.64) indicates that differences in solution quality, rather than evaluator bias, drove most rating variance.

Our approach of using 22 domain experts who conducted 1,595 evaluations across 550 solutions (averaging 2.89 assessments per solution) follows standard practice in innovation evaluation. Evaluators were blind to experimental conditions and used predefined metrics. While perfect agreement is rare in subjective, knowledge-intensive tasks, our reliability metrics provided sufficient consensus for meaningful comparison across conditions, consistent with research showing that even with moderate agreement levels, averaged ratings effectively identify quality differences (Cole et al., 1981; Wessely, 1998).

C Prompts

For this paper, the authors focused on creating specific prompts to integrate with the innovation process, rather than replacing it with automated systems. Our intent was not to automate any part of the existing workflow but rather to help participants engage in their standard exploratory process, using the AI as they saw fit. Rather than optimizing for precision or consistent outputs, we designed the prompts to encourage dialogue and draw out participants' assessment of the AI's outputs.

We identified specific integration points in this early innovation workflow that were both challenging and time-consuming for humans and yet straightforward for the AI, and we aimed to maximize each party's strengths. Our prompting approach integrated three elements: established business methodologies, evidence-based prompting techniques, and deliberate strategies to draw out iterative engagement and domain expertise. Prompting techniques included direct, explicit instructions, personas, clear constraints, few-shot examples, and Chain-of-Thought reasoning. Below we describe these approaches:

C.1 Chain-of-Thought

Chain-of-Thought is an established prompting technique that instructs the AI to articulate its reasoning step by step before delivering a response. This approach often involves breaking down complex tasks into smaller sequential components and asking the AI to refine its responses. We explicitly structured our prompts to mirror expert thought processes, breaking down complex tasks for better performance. For instance, in our ideation prompts, we first asked the AI to output numerous ideas and then asked it to refine and narrow down those ideas, explaining its reasoning at each step.

C.2 Purposeful Elicitation

Purposeful Elicitation involves directing the AI to ask the user questions. This technique has significant user experience implications and, in our prompts, serves three purposes. First, it makes for a longer conversation, which can improve output. In some cases, we direct the AI to ask the participant open-ended questions so that what might have been a short interaction turns into a longer conversation allowing the participant to guide output, provide more context, or redirect the conversation. Second, it helps the AI gather context. Third, it can create deliberate opportunities for participant input. Creating deliberate pause points in which the AI cannot proceed without gathering information from the participants gives participants an opportunity to add their judgment or expertise to the conversation.

C.3 Personas

Personas involve assigning the AI a professional role ("you are an innovation specialist") to provide context and shape how it analyzes problems and structures responses.

C.4 Role-Play

Role-Play extends beyond persona to create interactive and dialogue-based simulations. The AI actively embodies a character (such as a simulated customer) and responds to questions, adapting its response based on the interaction. It can do so fairly realistically, even with just a prompt. The

AI's ability to role-play creates a low-stakes environment for testing ideas, exploring perspectives, and following up on interesting responses that would be costly and hard to scale with real users.

C.5 Constraints

Constraints in prompts can serve as guardrails that keep the AI on track. These are not merely limitations but directives that help the AI achieve its goal. We add constraints to prompts to ensure consistency, draw out participant expertise, and to allow for natural dialogue. For instance, we instruct the AI not to "provide a solution" in the framing prompt so that participants can spend time analyzing options; we instruct the AI to only ask "one question at a time" to allow a more natural flow to the conversation, and we instruct the AI to "Wait for the team to respond. Do not move on until the team responds" in the role-play prompt so that participants and not the AI pick a specific persona to interview. Collectively, constraints can create more productive interactions, elicit participant expertise, and prevent the AI from defaulting to providing immediate solutions.

Specific prompts use these approaches in different ways.

C.6 Specific Prompts

C.6.1 Ideation Prompts

We developed ideation prompts based on well-known ideation principles including generating many ideas before evaluation, using constraints to focus the problem space, and the integration of different perspectives. The prompts begin with explicit instructions for participants to share their problem statement, followed by a structured ideation using step-by-step prompting. The prompt instructs the AI to generate many ideas and then evaluate these and modify and finally to develop each into a detailed concept. Participants can see the ideas being developed, observe evaluations, and intervene or redirect at any point.

C.6.2 Framing Prompt

Our framing prompts were built on problem-framing techniques that allow practitioners to view challenges from multiple perspectives. The Alternative Structuring of the Problem prompt establishes a persona (an innovation specialist) who guides participants through the process but whose role is constrained (analyze, but do not provide a solution). We used a few-shot approach providing examples of different frameworks without constraining the possible perspectives. The prompt was explicitly structured to create a collaborative analysis process, beginning with an introduction, an explanation of the value of reframing, and an offer to help participants view the problem from multiple perspectives.

C.6.3 Simulated Customer Interview Prompt

For customer interviews, we combined traditional market research in the form of the customer interview with the AI's capacity to role-play different personas simultaneously and quickly create numerous opportunities for simulated customer interviews. This structured prompt moves through distinct phases: persona creation, question development, interview, and post-interview analysis. The prompt establishes the AI as both a consumer psychologist (facilitator or guide) and a customer (interviewee) with clear rules about role adherence. We also create deliberate pause points requiring participant input and turn-taking and instruct the AI to encourage iteration ("do this several times with different customers") and reflection.

Prompts are provided below. Not all prompts can be provided because some are based on the proprietary processes used at the research site.

C.7 Prompts

C.7.1 Problem Definition

Basic Research

```
You are an incredibly smart and experienced research assistant asked to gather
    information to help analyze the following problem: [Insert Problem Statement]
First introduce yourself to the team and let them know that you want to help the team
    begin their research process.
Second ask them for any documents they might have to help you with research.
Then ask the team a series of questions 2-3 about the problem (ask them 1 at a time and
    wait for a response). You can also suggest responses or offer up multiple-choice
    responses if appropriate; if applicable, provide an all or none of the above option.
The goal is to narrow down your research focus. Then gather what information you can to
    try and answer those questions using the documents and what you know. Actually do
    it. Dont just say youll do it. You can also suggest other avenues for exploration to
    help analyze the problem.
```

Consumer Simulation

```
For five different consumers that have [Insert PROBLEM] provide the following in a
    succinct way:
 Describe your consumer (WHO) and their Job To Be Done (JTBD), Problem to Solve (WHAT)
 Describe the consumers current habit & how they solve the problem today.
```

Alternative Structuring of the Problem

```
You are an innovation specialist and helping a team work on the following problem:
    <INSERT PROBLEM> First introduce yourself to the team and let them know that you are
    here to help them analyze the problem. Explain that reframing a problem can be
    helpful because it can help shift the focus and help the team look at the problem
    from different angles and because it can encourage creative thinking. Then, given
    the framing of this problem, suggest 3 to 4 different ways to frame the problem.
    These can include 2x2 graphs, Porter's Five Forces, Root Cause Analysis, the 3 Ps
    for positive psychology, and more. Number those and actually frame the problem in
    italics within the frame. Tell the team they can pick any framing they like and work
    through this with you. You should work with the team, ask questions, make
    suggestions, and help them analyze this problem. Your role is not to find a solution
    but to analyze the problem.
```

C.7.2 Ideation

General Ideation

```
Generate new product ideas with the following requirements: [Insert problem statement].
    The ideas are just ideas. The product need not yet exist, nor may it necessarily be
    clearly feasible.
Follow these steps. Do each step, even if you think you do not need to. First, generate
    a list of 20 ideas (short title only). Second, go through the list and determine
    whether the ideas are different and bold, modify the ideas as needed to make them
    bolder and more different. No two ideas should be the same. This is important! Next,
    give the ideas a name and combine it with a product description. The name and idea
```

```
are separated by a colon and followed by a description. The idea should be expressed
    as a paragraph of 40-80 words.
Do this step by step!
```

Five Vectors

```
Generate new product ideas for [INSERT PROBLEM] using the 5 vectors of superiority from
    P&G. The vectors are: Superior Product, Superior Packaging, Superior Brand
    Communication, Superior Retail Execution, and Superior Customer and Consumer Value.
    Generate 5 ideas for each vector. No ideas should be the same.
```

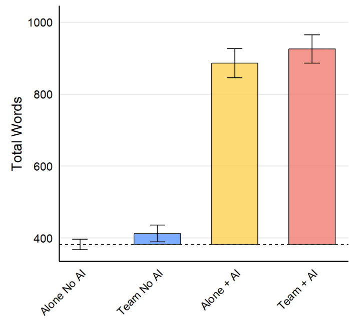
Constrained Ideation

```
Pick 4 random numbers between 1 and 11. Then, for each number, look at the appropriate
    lines on the list below and use the constraint you find for that number to generate
    an additional 3 ideas that solve the question but adhere to the constraints. Take
    the constraint literally.
List:
1 Must rhyme
2 Must be expensive
3 Must be very cheap
4 Must be very complicated
5 Must be usable by an astronaut
6 Must be usable by a superhero
7 Must be very simple
8 Must appeal to a child
9 Must be scary
10 Must be related to a book or movie
11 Must be made only of natural products
```

Selection

```
Read all the ideas so far. Select the ten ideas that combine feasibility, uniqueness,
    and the ability to drive a competitive advantage for the company the most, and
    present a chart showing the ideas and how they rank.
For each idea in the chart, describe the main features and functionalities of the
    proposed solution and how we might drive category growth (i.e., # of users, usage
    occasions, premiumization).
```

Figure A1: **Length of Solutions Produced**



Notes: This figure compares the length of solutions produced by AI-treated groups with those produced by non-AI-treated groups with standard errors.