Working Paper 26-011

The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders

Luca Vendraminelli
Matthew Dossantos Disorbo
Annika Hildebrandt
Edward McFowland III
Arvind Karunakaran
Iavor Bojinov

Harvard Business School

The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders

Luca Vendraminelli
Stanford University

Edward McFowland III
Harvard Business School

Matthew Dossantos Disorbo
Harvard Business School

Arvind Karunakaran
Stanford University

Annika Hildebrandt
Harvard Business School

Iavor Bojinov
Harvard Business School

Working Paper 26-011

Funding for this research was provided in part by Harvard Business School and The Stanford Digital Economy Lab with support from Google LLC.

The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders *

LUCA VENDRAMINELLI[1,2]
MATTHEW DOSSANTOS DISORBO[3]
ANNIKA HILDEBRANDT[3,4]
EDWARD MCFOWLAND III[3,4]
ARVIND KARUNAKARAN[1]
IAVOR BOJINOV[3,4]

[1]Stanford University

[2]Stanford Digital Economy Lab

[3]Harvard Business School

[4]Digital, Data, Design Institute at Harvard

Last update: September 8, 2025

*This study would not have been possible without the support of Olga Pirog, Ruby Lau, and Isaac Lai, who not only provided invaluable assistance throughout the entire process but also organized the experiment and collected the data used in our analysis. We would also like to acknowledge IG, the hosting company for this experiment, whose support and resources were indispensable to our research. The Stanford Digital Economy Lab was partially supported by a gift from Google LLC in the execution of this work.

The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders

ABSTRACT

As firms continue to seek efficiency by deploying Generative AI (GenAI) tools across various organizational functions, a critical question emerges: When and why does GenAI enable (versus constrain) individuals from one occupational group (i.e., outsiders) to perform tasks assigned to another occupational group (i.e., insiders), with equivalent speed and quality? And does GenAI's effect diminish as the "knowledge distance" between occupational groups increases? In an experiment conducted at a large UK firm, we examine these questions. Three groups—occupational insiders, adjacent outsiders, and distant outsiders—attempted to both conceptualize as well as execute tasks that are "core" to occupational insiders and were randomly assigned to receive support from a bespoke GenAI tool. We found that a "GenAI wall"—that is, the point at which GenAI can no longer meaningfully reduce the expertise gap between occupational insiders and outsiders—emerged for the joint effect of knowledge distance and task characteristics. Specifically, we found that GenAI is more effective at bridging expertise gaps between near (rather than distant) occupations, and more so for conceptualization (as opposed to execution) tasks. We discuss the implications of these findings for scholarship on occupations, learning, and division of labor in the wake of emerging technologies such as GenAI.

Keywords: GenAI, human expertise, labor demand, randomized experiment

The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders

INTRODUCTION

Scholars have predicted that Generative AI (GenAI) has the potential to transform many of the tasks that organizational members need to perform to deliver products and services (Jia et al., 2024). Eloundou et al. (2024) estimate that GenAI could transform at least 10% of job tasks for roughly 80% of the U.S. workforce, with about 19% of workers experiencing fundamental changes in at least half of their tasks. Moreover, GenAI is transforming not only the structure and execution of tasks, but also the underlying expertise and skills required to perform them (Yue et al., 2022; Autor, 2024).

Recent research has provided empirical evidence of task and expertise transformations as a result of GenAI's introduction into the work of various occupations, including writers (Doshi and Hauser, 2023), strategic consultants (Dell'Acqua et al., 2023), entrepreneurs (Otis et al., 2025), and call center agents (Jia et al., 2024; Brynjolfsson et al., 2024). However, most of the current research has focused on the effect of GenAI on the divide between low- and high-performing individuals within the *same* occupation. Some of these studies suggest that the gap between high and low performers within a single occupation is widening, as high performers benefit more from GenAI than low performers (Otis et al., 2025). Other studies point to the opposite pattern, arguing that GenAI can absorb the expertise gap and act as a "skill-leveler" by helping lower performers catch up with higher performers (Brynjolfsson et al., 2024; Dell'Acqua et al., 2023).

Although these studies have improved our understanding of the impact of GenAI on the performance of low and high performers *within* the same occupation, it is unclear how GenAI could potentially impact performance *between* different occupations, and especially between organizational members who are occupational insiders versus outsiders. If GenAI could reduce the expertise gap between occupational insiders and outsiders, this would imply that outsiders from entirely different occupations (and organizational functions) within the same firm might be able to assume roles that once required occupational insiders with extensive training and credentials. For example, data scientists could transition into another function and occupation within the same organization (e.g., marketing analyst, financial analyst) with significantly less retraining than would otherwise be needed without GenAI. But there is limited empirical evidence on these effects and their scope conditions. Therefore, in this study, we ask: When and why does GenAI equip (versus constrain) occupational outsiders to perform tasks at a equivalent level of productivity and quality of occupational insiders? And does GenAI's effect on the divide between outsiders and insiders diminish as the "knowledge distance" between occupational groups increases?

Our hypotheses are grounded in the theory of transfer of learning, which posits that an individual will master an unfamiliar task more effectively when it shares features with tasks they already perform well (Singley and Anderson, 1989). We apply this theory to a context where occupational insiders, adjacent outsiders, and distant outsiders (each having a distinct "knowledge distance" from insiders) perform tasks predominantly performed by insiders. We define knowledge distance (between insiders and outsiders) as the measure of dissimilarity between the skill demands of their job tasks. Simply put, a smaller knowledge distance—such as that between insiders and adjacent outsiders—indicates a greater overlap of the underlying skills required to execute their respective job tasks. In contrast, a wider knowledge distance—such as that between insiders and distant outsiders—indicates lower skill overlap. Drawing on the theory of transfer of learning, we propose that the ability of individuals to effectively leverage GenAI (i.e, in a way that emulates the performance of insiders) for tasks declines as this knowledge distance increases. Consequently, we hypothesize that GenAI will enable adjacent outsiders to close their performance gap with insiders more significantly than distant outsiders. In other words, we predict a "GenAI wall effect" for distant outsiders (as opposed to adjacent outsiders)—that is, the emergence of a point at which GenAI can no longer meaningfully reduce the expertise gaps between insiders and outsiders because of the wider knowledge distance between their jobs.

We tested our hypotheses using data from a multi-method study in collaboration with IG, a large global trading company headquartered in the UK. We advised the company's AI team in the design of a randomized experiment to test this hypothesis [1]. The company collected data from 78 IG employees, each of whom participated in an experiment requiring them to write two web articles for the company's website. These 78 employees were selected from three distinct groups: 12 web analysts, 26 marketing specialists, and 40 technology specialists. The participants were classified into these respective groups based on the nature of their job responsibilities and the organizational function they belong to. The web analysts are primarily responsible for writing web articles for the IG website, so they are categorized as *occupational insiders* (hereafter, "insiders"), as the task was part of their current jobs. Marketing specialists work in the same marketing function as web analysts, where they undertake marketing-related tasks but typically do not write web articles; therefore, they are classified as *adjacent outsiders*, for whom writing web articles is unfamiliar but still aligns with their core marketing tasks. Finally, technology specialists, mostly software developers and data scientists from the technology operations department, neither write web articles nor engage in marketing-focused activities. As a result of this knowledge distance, they are classified as *distant outsiders*, since writing marketing articles is an unfamiliar and unrelated task to their primary job responsibilities.

By analyzing the experimental data collected by the company, we aimed to determine whether marketing specialists and/or technology specialists could develop web articles at a level comparable to web analysts, and whether GenAI would facilitate this achievement. In the experiment, all participants followed the same procedure that web analysts typically used to create two articles: first, they laid out the conceptual framework of each article (articulating the main points and structure), and then they transformed these conceptual outlines into final, fully developed articles. This allowed us to test the difference between groups' outputs in both the conceptualization and execution phases. Access to the bespoke GenAI tools to support article conceptualization and creation was randomized as described in Section 3.

We have access to multiple types of data collected before, during, and after the

[1] The project has received IRB approval, IRB24-1269. The experimental design was preregistered at `https://aspredicted.org/hx4b-mtsw.pdf`.

experiment. Prior to the experiment, one of the authors collected 10 semi-structured interviews with web analysts, marketing specialists, and technology specialists [2]. During the experiment, the company collected individual task completion times, system interaction logs, and participants' final outputs. A team of MBA students, and two of the authors of this paper, then evaluated each participant's final output using a rubric created by the head of web analytics at IG, who was also responsible for approving article publications on the corporate website. Following the experiment, one of the authors conducted 18 interviews with the experiment participants to better understand how the three occupational groups approached the experimental tasks. The company also administered pre- and post-experimental surveys capturing demographic information, personal attitudes, GenAI usage behaviors, writing styles, job design details, and IT/GenAI expertise. We use data from these surveys and tests as controls for individual differences in our analyses.

Our analysis of the impact of GenAI on the performance of marketing specialists, technologists, and web analysts in article conceptualization and execution tasks reveals two main insights. First, we find a differential effect of GenAI for marketing versus technology specialists in task conceptualization rather than execution. In task conceptualization (e.g., conceptualizing the article), we identified low-performing marketing and technology specialists who were unable to match the quality of web analysts without the support of GenAI. However, the introduction of GenAI has leveled the playing field by allowing lower-performing marketing and technology specialists to achieve performance levels comparable to those of web analysts (in terms of quality and speed). In the article execution,[3] the patterns diverged: for technology specialists, GenAI did not alter the performance distribution, leaving a long tail that was unable to close the gap with web analysts. By contrast, GenAI shifted the quality output distribution up for marketing specialists, allowing them to bridge that gap with web analysts.

Second, the analysis of the post-experimental interview suggests a potential explanation for why only marketing specialists were able to close the divide with web analysts for article execution, grounded in the interplay between the knowledge distance from insiders and the task

[2]The interviews conducted throughout the research project were covered by IRB-72148.

[3]Due to data limitations, we could not draw firm conclusions on performance differences without GenAI.

characteristics. On the one hand, we find that the three groups had differential approaches to task resolution due to their different expertise, and our qualitative results show that this “knowledge distance” was critical in determining such differences in task resolution behavior. Specifically, the task resolution approach of marketing specialists was far more similar to that of web analysts than to that of technology specialists. On the other hand, the qualitative interviews also suggest that the impact of this differential approach varied from conceptualization to execution because the two tasks demand fundamentally different outputs. Conceptualization was an act of abstract formulation, which required listing the core features and ideas for an article. For this task, GenAI’s support in defining features, combined with basic intuition, was largely sufficient, making the different approaches less consequential. In contrast, execution was an act of embodiment, which required elaborating on the listed features and ideas and giving them a tangible form and structure through prose. Benefiting from the support of GenAI in this latter task was far more sensitive to the task resolution approach adopted; this is akin to the findings in Dell’Acqua et al. (2023), which identified a jagged technological frontier. Technology specialists lacked the fundamental rudiments of marketing content creation, and this prevented them from exploiting their GenAI article recommendations and closing the quality gap with web analysts.

Together, these findings articulate a mechanism through which GenAI equips (versus constrains) adjacent or distant outsiders (i.e., marketing and technology specialists) to perform tasks with the same speed and quality as individuals in another occupational group (i.e., web analysts). Our results reveal the potential presence of a “GenAI wall”: a point at which GenAI can no longer meaningfully reduce the expertise gap between insiders and outsiders, because sufficiently “distant” outsiders lack the foundational knowledge required to utilize GenAI recommendations effectively. Indeed, this wall delineates the zone where AI can effectively bridge expertise gaps from the zone where deep domain knowledge remains indispensable. We find that the location of this wall depends on the interplay between the characteristics of the task and the knowledge distance from insiders. Specifically, for conceptualization tasks, which center on high-level ideation, the GenAI wall seems not to exist—or rather GenAI alone may be able to perform at the level of experts. Indeed, we found that both adjacent outsiders (i.e., marketing specialists) and distant outsiders (i.e., technologists) match the performance of insiders (i.e., web analysts). However, for the execution task—which shifted the focus from ideation to the detailed crafting and refinement of the article—the wall emerged for distant outsiders (i.e., technologists), whose divergent work processes impeded effective AI use.

Our results contribute to our understanding of the impact of GenAI's on expertise by distinguishing between vertical and horizontal performance convergence. Much of the current research highlights how GenAI compresses performance variation by benefiting lower-skilled individuals, thus shrinking the gap between low- and high-performers within a single occupation (Doshi and Hauser, 2023; Dell'Acqua et al., 2023; Brynjolfsson et al., 2024). We conceptualize this phenomenon as the "GenAI ceiling"—the vertical limit of AI's ability to elevate the output of low-performers toward the levels of high-performers within a given occupation. This focus, however, overlooks a more disruptive, horizontal dynamic: how GenAI facilitates cross-occupational encroachment. We address this gap by introducing a complementary theoretical concept: the "GenAI wall." This wall represents the horizontal limit of AI, where the knowledge distance between an outsider and an insider—for specific tasks—becomes too wide for GenAI to bridge effectively.

Our work also speaks to the rising literature on AI-assisted creative work that focuses on the "Ideation-Execution gap" (Si et al., 2025). This literature, so far, has focused on how the quality of AI-generated concepts can degrade during practical implementation. We update this view by shifting the focus from the conceptualization act to the execution act. While previous literature focused on the consequences of poor conceptualization on execution, we argue instead that the act of execution is itself a much more complex task than conceptualization because it requires embodiment rather than abstract conceptualization. Controlling for idea generation quality, we find that when knowledge distance increases, GenAI struggles more with the hands-on work of making ideas tangible and concrete than with supporting abstract conceptualization.

Finally, our study offers task-level evidence that may open new avenues of research to challenge the foundational theories of specialization and the division of labor. We show that GenAI shifts task performance away from tacit knowledge acquired through deep, task-specific practice (Dreyfus and Dreyfus, 2005; Ericsson and Charness, 1994) toward broader 'foundational knowledge' that enables effective human–GenAI collaboration. This insight carries two theoretical implications. First, it challenges conventional models of expertise by revealing an alternative pathway to high performance not predicated on hands-on practice (DiBenigno, 2018; Rousseau and Stouten, 2025). Second, it opens new directions for research on the division of labor. Foundational organizational theory has long held that high knowledge-transfer costs necessitate bundling similar tasks into jobs and maintaining their stability over time, enabling individuals to progress along learning curves (Smith, 1776; Simon, 1945; Hayek, 2013). Our findings call this assumption into question, suggesting that GenAI can significantly reduce the knowledge barriers that have historically constrained work design (Cohen and Bui, 2024; Postrel, 2002) (in our case, GenAI enable marketing specialists to do the job of web analyst without retraining). This raises important questions about whether, and under what conditions, firms will leverage GenAI to expand task variety and dismantle traditional job silos.

THEORETICAL BACKGROUND

Past research on technology and organizations has shown that the diffusion of new technologies can reshape both the tasks that workers within organizations perform and the expertise required to complete them (Autor and Handel, 2013; Yue et al., 2022; Autor, 2024). For instance, during the first industrial revolution, manufacturing shifted from relying on skilled craftsmen to employing factory workers. As machines replaced traditional craftsmanship, workers were required to learn how to control and maintain engines instead of focusing on artisanal tasks. Similarly, the advent of computers shifted the demand away from handwriting and manual data entry toward software-related expertise (Agrawal et al., 2018). In both cases, the introduction of new technologies altered the tasks necessary for organizations to deliver value to customers and, consequently, changed the expertise that workers needed to be relevant within organizations.

Individuals previously suited for craft or clerical work had to develop new skills, becoming proficient at operating machines or using computers to produce and manage documents (Autor, 2024).

Like these historical examples, GenAI also has the potential to alter many of the tasks workers within companies perform and the expertise required to perform them. Researchers have demonstrated that GenAI—by informing, recommending, or deciding—will affect the productivity and quality of the tasks performed by many occupations working within organizations. For instance, in professional writing, Noy and Zhang (2023) reported that ChatGPT usage reduced task completion time by 40% and improved output quality by 18%. Similarly, in customer service, Brynjolfsson et al. (2024) found that GPT-based support led to a 14% increase in the number of customer issues resolved per hour. Meanwhile, in strategic management consulting, Dell'Acqua et al. (2023) observed a 46.6% performance boost, although the extent of the gains varied by task characteristics. Collectively, these findings suggest that GenAI transforms how tasks are accomplished, thereby influencing both individual productivity and the quality of outputs. Moreover, research has shown that such increases in productivity free up individuals' time, allowing them to change their occupational task composition (Hollister et al., 2023). For instance, Jia et al. (2024) found that in sales departments, organizations can use GenAI to offload routine and mundane tasks, enabling human workers to focus on more complex, high-value activities that demand creative problem-solving and innovation.

Prior research has also begun to explore how tasks transformed by GenAI are reshaping the expertise required to perform them effectively. Specifically, this research has focused on the "vertical" effects of GenAI, meaning how, within a single occupation, GenAI influences the divide between high- and low-performing individuals. However, the results so far have been conflicting. Some studies suggest that GenAI will increase the value of expertise, thus widening the divide between high- and low-performance performers, since high-performing performers are likely to reap greater benefits from GenAI than low-performing performers (Berger et al., 2024). In a field experiment, Otis et al. (2024) showed that Kenyan entrepreneurs who performed better with

GenAI were significantly more adept at using it, crafting better prompts and achieving superior results, while lower-performing entrepreneurs achieved smaller performance gains with GenAI. Conversely, other studies suggest GenAI narrows the gap between high and low performers by providing greater benefits to those who initially have lower expertise levels. Some research suggests that professionals will need less expertise to achieve the same level of performance, showing that lower performers benefit more from GenAI than high performers, with the gap between the two groups becoming statistically insignificant when GenAI is in the decision-making loop (Noy and Zhang, 2023; Brynjolfsson et al., 2024). For instance, Doshi and Hauser (2023) found that GenAI had a greater positive impact on less creative writers (as measured by the DAT test) than on those who were already highly creative. This result is specifically interesting because it, unlike the former results, indicates that GenAI will be able to absorb task expertise.

While existing studies offer valuable insights into how GenAI reshapes the value of expertise within specific occupations, their narrow focus on intra-occupational dynamics—such as narrowing performance gaps between high- and low-performing individuals—risks overlooking broader organizational implications. In particular, these studies have paid less attention to GenAI's "horizontal" effect, or rather its potential to reconfigure how organizations divide labor across specialists. Crucially, previous research has not examined whether GenAI could reduce the retraining needed for individuals to take on unfamiliar tasks under the jurisdiction of other occupations.

2.1 Hypotheses Development

We conceptualize an organization as a system of processes, each comprising tasks that demand specialized expertise (Hayek, 2013; Simon, 1945; Garicano, 2000). Within this system, employees who routinely perform a specific task (we call them insiders) develop deep, hands-on expertise through repeated engagement, earning them recognition as task experts. For example, in the fashion industry, "allocators" become insiders to the task of distributing merchandise from warehouses to stores, a role distinct from that of "sales associates," who are insiders to the task of advising customers on the shop floor.

For any given task, however, there are also organizational members who do not perform it regularly (we call them outsiders). Lacking the insider's context-specific, hands-on expertise, these outsiders face a knowledge barrier to perform that task with the quality of insiders (Dreyfus and Dreyfus, 2005; Ericsson and Charness, 1994). We hypothesize that GenAI can dismantle this knowledge barrier. By codifying and delivering task-relevant knowledge—informing, recommending, or even executing decisions—GenAI can substitute for the tacit expertise that outsiders lack. This capability allows GenAI to bridge the knowledge distance that has traditionally separated occupational roles. Therefore, we formalize the following hypothesis:

> *Hypothesis 1. GenAI reduces the performance divide between task insiders and outsiders.*

While 1 suggests that GenAI can help all outsiders, we argue that not all outsiders are equally distant from the insiders' knowledge. Within an organization, occupations can be close to or distant from one another depending on the similarity of the skills their job tasks require. [4].

This concept of knowledge distance from insiders is critical because of its implications for determining the extent to which outsiders can close the quality and speed gap with insiders. The theory of learning transfer suggests that an individual's ability to perform an unfamiliar new task is directly related to how much it overlaps with old tasks they have already mastered (in more formal terms, this "overlap" is the similarity between the skill-demand vectors of the new and old tasks). At its core, the theory posits that individuals approach unfamiliar tasks by redeploying knowledge accumulated from previous tasks, and, as such, this knowledge is more valuable when applied to similar rather than dissimilar ones. For example, expertise in playing the saxophone can be readily transferred to learning the flute, as both share similar fingerings and principles of breath control, making the transition relatively smooth. In contrast, saxophone expertise offers minimal

[4] We define this relationship within a high-dimensional *skills space* $\mathcal{S}$, whose N dimensions correspond to the set of skills $K = \{k_1, k_2, \ldots, k_N\}$ required to operate the organization. Each occupation O_i is assigned a set of tasks M_i, each of which is represented by a skill-requirement vector in $\mathbb{R}^N$. Aggregating across tasks produces the occupation's overall skill-demand vector $\mathbf{v}_i \in \mathbb{R}^N$. The *knowledge distance* between two occupations O_i and O_j is the dissimilarity between their skill-demand vectors. Smaller values of $d(O_i, O_j)$ indicate lower knowledge distance.

advantage for a “distant” task like baking a cake or performing surgery.

This principle explains why we differentiate between “adjacent outsiders,” for whom the focal task is comparatively closer in knowledge distance, and “distant outsiders,” for whom the task is significantly farther away. The theory of learning transfer pushes us to predict that closing the divide with insiders is fundamentally different for adjacent and distant outsiders. Adjacent outsiders engage in *near transfer*—transferring knowledge from tasks that require similar skills—while distant outsiders face the more cognitively demanding challenge of *far transfer*, which involves adapting knowledge from tasks requiring a different set of skills. Although GenAI can serve as support for both, we argue its affordances are better suited for near transfer than for guiding the deeper knowledge realignment essential for far transfer. In other words, we suggest that GenAI’s utility diminishes as knowledge distance increases, eventually reaching a point where the divergence in proprietary knowledge becomes too large for the technology to reconcile. We predict a “GenAI wall effect” for distant outsiders (as opposed to adjacent outsiders)—that is, the emergence of a point at which GenAI can no longer meaningfully reduce the expertise gaps between insiders and distant outsiders. As such, we hypothesize:

> *Hypothesis 2. In a scenario where both adjacent and distant outsiders aim to match insider performance, GenAI will enable adjacent outsiders to reach performance levels closer to those of insiders than distant outsiders.*

We test 1 and 2 with the experimental data described below.

METHODOLOGY

This research was conducted within IG, a leading fintech company that operates a global trading platform. At the time of the study, IG had a robust user base of more than 350,000 active users, generated approximately 1 billion in annual revenue, and employed more than 2,000 employees worldwide. Primarily a trading provider platform, IG allows investors to execute trades across various financial markets using its advanced technology. The company invests heavily in product development, supported by a dedicated team of developers and data scientists, and strategic marketing, which is managed by the marketing department. The experiment was conducted primarily within these two departments.

Our first site visit took place in early 2024, when the last author initiated a partnership during a visit to IG's headquarters in London, UK. This marked the beginning of a series of in-depth meetings between our teams to explore IG was planning to realease GenAI into its operations. In these early meetings, we learned that, following a company reorganization, leadership tasked the global head of data and AI transformation and her team ("AI team") with creating new AI tools and driving their adoption within the organization. Our team has been working closely with the AI team ever since.

At the time we established our partnership with IG, the AI team had spent the last few months developing a GenAI tool for creating online content; the tool was ready for use and slated for release in late 2024. This tool was powered by Large Language Models (LLMs), fine-tuned with documents from a recent collaboration with a marketing agency that had helped IG define its corporate style and content requirements. The AI team was confident that the output from the tool would align well with their brand identity and maintain a consistent communication style more effectively than the other generic LLM models available on the market. We decided to investigate the potential impact this bespoke tool could have within the organization.

The IG team identified the internal SEO department as the first use case to test the effectiveness of this LLM-based tool for content creation. At the time of our exploration, the SEO department, consisting of 10 members (plus external contractors and the SEO team leader), was tasked with overseeing the web strategy and content production for the IG website. One of the SEO team's primary responsibilities is analyzing web trends to identify and produce content that, if published on the IG website, would drive user engagement, specifically for IG's consumer segments. The AI team believed that the SEO team would benefit greatly from GenAI, which would help web analysts produce high-quality web article content more efficiently. As such, we planned a series of interviews and meetings with key stakeholders to better understand the structure and daily tasks of the SEO team and confirm their hypothesis.

Through our interviews with web analysts, we discovered that writing web articles was one of their core occupational tasks. In particular, web analysts told us that, on average, 20-40% of their time was dedicated to writing web articles. Article quality was crucial, especially in the context of evergreen content (designed to remain on the website indefinitely) to rank IG's website at the top of Google searches and guarantee customer engagement. This engagement allowed them to drive more traffic to the site and, ultimately, convert it into sales. When we looked more closely at the creation of evergreen content, we found that the process produced about 60 articles annually.

From our interviews and review of the literature, we concluded that adopting GenAI could alter the article-writing process and its tasks. However, its impact on productivity, output quality, and, more broadly, changes in the expertise required to perform web article writing was unclear. In order to study this open question, we advised the company on the design of an experiment in which three occupational groups—web analysts, marketing specialists, and technology specialists—created web articles both with and without GenAI. IG decided to carry out this experiment and provided us with the data, enabling us to write this manuscript.

3.1 Occupational Groups

For this experiment, the AI team recruited employees within IG and asked them to conceptualize and write two web articles for the corporate website. The first group of participants was categorized as insiders and consisted of the 12 web analysts who regularly wrote web articles for the corporate website. The other participants were classified in two distinct groups based on their knowledge distance from web analysts. We identify 24 marketing specialists, who worked in the same department of web analysts and did similar marketing-related tasks, as adjacent outsiders. We classify 40 technology specialists, who worked on the technical department and reported to a different C-level manager, as distant outsiders. Below, we describe the rationale with which this classification was made, and in Table 1 we present descriptive statistics about the three different groups.

3.1.1 Web Analysts: Insiders. The first group of participants includes all web analysts, who were primarily responsible for creating web articles within the company. Creating article conceptualizations and executing articles occupied approximately 20-40% of web analysts' time. As one web analyst described her job:

> "Our team is essentially responsible for acquiring organic website traffic for our company's website. This includes overseeing SEO [search engine optimization] and CRO [conversion rate optimization]. We focus on how people visit our site through Google, their journey after arriving, and how they come back. We generate web articles to increase traffic and edit the website, but that only accounts for 20-40% of our time. Everything we do is to improve our web performance." (Web Analysts - TE_04)

The web analysts that participated in the experiment had an average tenure of more than 2 years in their role. Additionally, all web analysts possessed a marketing background. As such, we defined this group as *insiders*.

3.1.2 Marketing Specialists: Adjacent Outsiders. The second group includes various IG employees working in the marketing department who were not web analysts. As such, writing web articles was not their main occupational task, but they had extensive experience with marketing-related tasks. For example, one of the participants described his role as digital marketing specialist as:

> "I collaborate with channel specialists to manage campaigns—whether using display programmatic ads, paid search, or other paid media channels." (Marketing Campaign Specialist - DE_03)

Another described her role as CRM marketing executive as:

> "As a CRM marketing executive, I focus predominantly on client communication and designing customer journeys. For instance, when the product team informs us of a pricing change for one of our products, I simplify the information for our clients and create communications to inform them of the update." (CRM Executive - DE_04)

Although employees in this group did not write articles for the IG website and were therefore unfamiliar with the task, they described core marketing responsibilities such as campaign planning, public relations, and YouTube content creation. These tasks were core marketing field responsibilities, making their roles adjacent to those of web analysts. Therefore, we will refer to this group as *adjacent outsiders*.

3.1.3 Technology Specialists: Distant Outsiders. This group included employees that were not part of the marketing department. Mostly, these employees belonged to technical departments, and their main job tasks were related to coding. For example, one technical specialist said:

> "I'm essentially a math modeler—but you could call me a data scientist. We work on things like recommender systems and, more recently, LLMs. There is a lot of work involved in training them with data" (Data Scientists - NO_01)

Another said:

> "I'm a software developer, and I mostly work on the front end and dashboards. My tasks are fairly common, like adding buttons here and there. Actually, I just added some feedback buttons right before coming here" (Software Developer - NO_02)

Another said:

> "I work in the data science group, but I'm a data engineer. I primarily transfer data from one location to another to enable others to develop models. For instance, we might have a client-value model that predicts how much value a client can generate for us. This requires pulling all the data we have about that customer into one place, which is not always a simple task. Anyhow, that's my job." (Data Scientist - NO_03)

The final group consists of employees from the technical department, with roles ranging from building mathematical models to developing software and maintaining data structures. Like the marketing specialists, technology specialists do not write web articles in their jobs. However, a critical distinction sets them apart: while marketing specialists operate within the same broad functional domain as marketing content creation, these employees' daily work involves coding and data modeling—tasks that are functionally and cognitively distant from marketing communications. This greater knowledge distance from web analysts led us to classify them as *distant outsiders*.

3.2 Experimental Design

3.2.1 Experimental tasks. During our experiment's design phase, we conducted nine interviews with web analysts to understand the article-writing process at IG. We found that the task generally involves two sequential tasks: creating an article conceptualization and then writing the article itself. First, web analysts produced an article conceptualization using a fixed template. This template included essential SEO and CRO elements such as keywords, URLs, meta-descriptions, headings (H1 and H2), FAQs, and suggestions for internal links. The conceptualization process usually started with a customer needs analysis, and the resulting conceptualization is reviewed by the head of the Web Analytics. Once approved, the conceptualization moves to the article execution task, which is essentially a free-form exercise starting from a blank page. Web analysts still respected certain requirements—such as adhering to a specific length, following the website's style guidelines for headings, and incorporating all the planned information from the article conceptualization. Although the article conceptualization sets the content and structure, authors retain considerable creative freedom to craft the final text and tailor it to the target audience.

Based on these insights, we proposed an experimental design to the company in which each participant was asked to create two article conceptualizations and write two articles using those conceptualizations. The conceptualization and article content were both related to IG's core business but addressed different topics: one focused on trading gold futures, emphasizing the advantages of pre-determining the purchase price and time for gold, and the other covered trading commodity CFDs, focusing on the potential benefits of speculating on commodity price fluctuations. These topics were selected and prepared by the head of the IG SEO team—who was otherwise not involved in the experiment—in collaboration with the AI team and our research team. Importantly, the topics reflected past IG web article themes and aligned closely with the content adjacent outsiders might handle in their day-to-day work.

Before starting the experiment, participants received a summary of the activities they would perform during the experiment, including detailed instructions on how to perform each task. The employees were given the same template used by the IG web analysts to help them create their article conceptualization. This template was the standard format established by the SEO team over time, and included fields for keywords, article URL, meta description, and primary and secondary headings (H1 and H2). Employees were not given a template for composing the web article, but were given an example of a web article that had been published on IG's website. Additionally, employees were instructed to ensure that the articles were written in plain, high-quality English, were SEO and CRO-optimized, and included internal links to related content on the company website.

3.2.2 GenAI tool and how it changed the execution of experimental tasks. In the treatment, tasks were performed with the support of two bespoke and fine-tuned LLMs. These models were developed and designed specifically by IG for article conceptualizations and web articles, respectively, and were deployed on internally accessible platforms.

The article conceptualization interface allowed the employee to enter the article topic as a prompt. The employee needed to specify the configuration parameters, including the specific LLM to be used (i.e., the fine-tuned version of ChatGPT, Claude, or Gemini), as well as the number of suggested keywords, URLs, meta-descriptions, H1 subsections, H2 subsections, FAQs, and internal link suggestions. After submitting the prompt and parameters, the employee would receive a completed conceptualization. For the second task, the interface used for article writing allowed the employee to upload an article conceptualization. After prompting the tool using the conceptualization, the employee would see a fully generated article based on the brief, which they had the option to download.

In both tasks, employees were still responsible for both prompting the LLM and refining the LLM-generated content—whether it was the article conceptualization or the final article—before submitting their work. This editing step was crucial: for instance, the conceptualization tool frequently provides multiple suggestions for fields such as URLs, meta-descriptions, and headings, requiring the human participant to select the most appropriate one. Without careful edits, the resulting output may have included superfluous or confusing information.

To ensure understanding of tool usage, employees were provided with a video explaining how to use the tool to generate the desired content prior to completion of the task. The video explained in detail the edits that the employee would need to make to the AI-generated output prior to submission, such as removing additional, unneeded fields.

3.2.3 Recruitment and incentives to participate. Across these three groups, 114 employees were invited to participate in the experiment. Most of the incentives to participate in the experiment came from IG's leadership, who promoted the event internally. First, the IG leadership encouraged team leaders to increase registration rates with internal communications that supported this initiative. Second, participating employees received a £20 Amazon voucher for registering for the experiment. Finally, ten £100 Amazon vouchers and ten one-year ChatGPT Pro licenses, were randomly distributed to participating employees. Following these incentives, 87 employees (76%) signed up to participate in the experiment. In total, 76 of the 87 submitted content for both the conceptualization and article execution tasks.

3.2.4 Randomization Strategy. IG randomized the task instructions regarding the use of the bespoke GenAI tools to support article conceptualization and article execution as follows. Participants completed two article conceptualizations on two topics (gold futures and commodity contract for differences (CFDs)) assigned in random order. All participants completed the first conceptualization without the use of GenAI, and the second conceptualization with GenAI. Participants used these conceptualizations as the basis for writing articles in the next task. For the full article execution, participants were randomly split into two groups: one group was instructed to compose both articles without GenAI, while the other was instructed to compose both articles with GenAI. The order in which the two articles were written in the second task was randomized, similar to the first task.

We randomized the order in which topics were completed to control for potential learning spillovers. Moreover, this randomization strategy addressed the challenge of monitoring employee compliance. Once employees were introduced to the GenAI tool, there was no mechanism to ensure that they would refrain from using it for subsequent tasks. All randomizations were

TABLE 1
Summary statistics for different expertise groups. Two employees in the "Adjacent Outsiders" group completed the article conceptualization task but not the article execution; therefore, we exclude them from this table.

	Insiders	Adjacent Outsiders	Distant Outsiders
Role, Department	SEO, Marketing	Non-SEO, Marketing	Non-SEO, Non-Marketing
Sample Size	12	24	40
Percent Female	33.33%	62.50%	37.50%
Percent Bachelors Degree or Higher	100.00%	91.67%	95.00%
Average Age	34.75 (5.56)	35.17 (7.33)	30.18 (6.51)
Average Professional Tenure (Years)	9.61 (5.23)	10.34 (7.35)	8.41 (8.63)
Average IG Tenure (Years)	3.37 (1.99)	3.14 (2.85)	2.82 (2.59)
Average Role Tenure (Years)	2.11 (1.02)	1.22 (0.97)	1.2 (1.95)
GenAI Anxiety Score (1-7)	2.73 (1.58)	2.78 (1.29)	2.53 (1.03)
GenAI Usage	3.42 (1.0)	2.58 (1.5)	2.98 (1.27)
Technological Change Attitude (1-3)	2.83 (0.33)	2.65 (0.58)	2.72 (0.42)
Career Attitude (1-7)	3.58 (0.95)	3.89 (1.25)	4.06 (1.32)
Job Satisfaction (1-7)	5.35 (1.3)	4.93 (0.85)	5.08 (1.03)
Co-worker Support (1-7)	5.73 (0.82)	5.46 (1.4)	5.6 (1.02)
Supervisor Support (1-7)	4.83 (1.72)	5.46 (1.41)	5.51 (1.29)
Information Processing (1-7)	6.08 (0.56)	6.02 (1.11)	5.94 (0.79)
Task Variety (1-7)	5.77 (1.02)	5.69 (1.51)	5.41 (1.26)
Schedule Autonomy (1-7)	5.56 (1.32)	5.25 (1.61)	5.58 (0.94)
Decision Autonomy (1-7)	5.64 (1.23)	5.36 (1.2)	5.59 (0.91)
Job Complexity (1-7)	2.31 (1.02)	2.52 (1.13)	2.54 (1.23)

executed to balance the three occupational groups. In total, as summarized in Figure 1, there were eight possible combinations of conditions to which employees could have been assigned.

For the article conceptualization task, 38 employees started with the gold futures conceptualization and 38 participants started with the commodity CFDs conceptualization.

For the article execution task, 37 employees started with the gold futures topic and 39 started with the commodity CFD topic. Additionally, access to GenAI was randomized. 47 employees were assigned to complete both articles with access to GenAI and 29 to complete both articles without it[5]. This randomization strategy — providing access to GenAI for either both articles, or neither of the articles — again helped ensure compliance by eliminating the risk of participants using GenAI for the second article if they had only been assigned GenAI use for the

[5]Note, in total, 76 employees submitted content for the article execution task. There were 11 people who completed the pre-survey and were randomized into treatment groups. However, they are not reported in this analysis as they did not submit both experimental tasks.

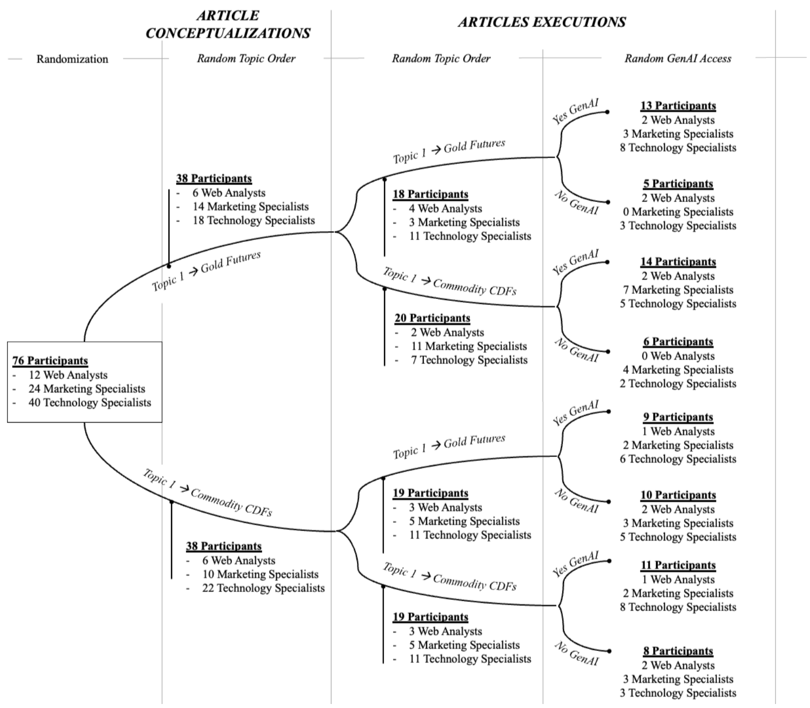

FIGURE 1
Employee randomization across both experimental tasks (creating conceptualizations and writing articles). Note, the above figure only includes employees that completed both the article conceptualization and web article tasks. There were two employees who completed just the conceptualization but who did not complete the web article.

first article. The sample sizes for each group are shown in Table 1.

IG provided our team with employee-level prompt log data from these tools during the experimental window (participants were unaware of our access to these logs). Unfortunately, we discovered noncompliance, finding evidence that employees prompted IG' GenAI tools despite being assigned to the 'Human Only' group for that particular task. Most notably, thirteen employees prompted the GenAI tool on both briefs (which we can identify based on the topics of the two briefs), even though employees were instructed to prompt the tool on just one brief. Further, four employees prompted the GenAI tool on articles despite being assigned to the Human Only group. Three employees used GenAI for both article tasks, while one employee used GenAI for one of the article tasks.

The number of employees that completed each task in compliance with instructions are reported in Figure 2. Figure 1 shows the original treatment assignments for employees that

	Article Conceptualizations		Article Executions	
	Human Only	GenAI	Human Only	GenAI
Web Analysts	9 (3)	9 (3)	3 (3)	6
Marketing Specialists	22 (2)	22 (2)	10	14
Technology Specialists	32 (8)	32 (8)	11 (1)	27

FIGURE 2
Completion numbers for each task with non-compliance. The number of employees who were dropped from analysis due to non-compliance are in parenthesis when applicable. Note, since each employee completed an article conceptualization with and without access to GenAI, the counts are the same in each treatment condition. Conversely, for article executions, employees were assigned to either complete both articles with GenAI or without GenAI. Therefore, value counts differ for each condition for article executions.

completed each task prior to non-compliers being removed.

It is possible these employees that did not comply with assigned treatment instructions were merely testing the GenAI tools, and did not actually use the output for their briefs or articles. However, to ensure that our results are as robust as possible, we removed these employees from the analyses of the respective task that they were noncompliant on.

We do not anticipate that this will impact the measurement of the treatment effect. For briefs, for instance, the thirteen excluded employees are 'always-takers': they used GenAI for both briefs, instead of just one brief as intended. Therefore, their treatment effect—the difference in quality, time or productivity between their briefs created with and without GenAI—is zero.

These exclusions were not in our pre-registration, since we did not anticipate any noncompliance. However, our main results—web analysts deliver the highest average quality across articles and briefs, combined, without GenAI, and GenAI closes the quality gap, especially for marketing specialists—still hold when we include these employees.

The experimental design and analysis was pre-registered[6] on 9/13/2024 and was launched a week later. The final employee to complete the experiment submitted their materials on

[6]https://aspredicted.org/hx4b-mtsw.pdf

1/23/2025. Ultimately, 78 of the 87 employees completed the first task (creating an article conceptualization), while 76 of those 78 employees completed the second task (writing an article). According to the AI team, the 12 web analysts comprise nearly all of the employees responsible for content creation. Any deviations from the pre-registration are discussed in Appendix B.

3.3 Data

In this research, we employed a mixed-method approach. Our findings come from the combined analysis of both qualitative and quantitative data gathered through field observations, interviews, surveys, and experiments. The following sections review the data we used and how they were collected.

3.3.1 Field Observations and interviews. Throughout our research, one of the authors conducted 10 pre-experimental interviews with web analysts, marketing specialists, technology specialists, and AI developers, shown in Table B1. After the experiment, the author conducted 22 post-experimental interviews with the same groups to understand their behaviors during the study. Additionally, we held weekly meetings with the corporate AI development team to review our findings and share impressions about the experiment. We used these interviews and observational data extensively to understand both the technology and the web article development process, to codify the expertise of the three main occupations involved, and to assess how each occupation approached the experimental task.

3.3.2 Experimental Data. We collected data on output quality and time to complete the experimental tasks.

The time to complete the task is a self-reported measure. After the completion of each task, employees were asked to submit their completed conceptualizations or final articles. They self-reported the time it took them to complete the task in minutes, as well as their percentage completion.

Assessing the content quality involved two steps: creating the evaluation instrument and measuring the quality of the experimental outputs. To create the evaluation instrument, we worked with the head of web analytics at IG. Within that organization, he was responsible for approving conceptualizations and accepting articles to be published on the corporate website. He created a detailed rubric for evaluating content, which can be found in the supplementary materials B.

The conceptualization and article grade is the average between all scores assigned by a grader. We recruited two types of graders for this research. First, we hired MBA students to evaluate each conceptualization and article. Before grading, we anonymized all materials by suppressing any identifying information. The MBA graders received only the details necessary for generating grades; they were unaware of the employees' different levels of expertise or that some employees wrote conceptualizations and articles with GenAI assistance. Second, two members of the research team independently graded the anonymized conceptualizations and articles. They then compared their assessments and, if their initial scores differed by two or more points, they discussed and potentially revised their final grades. For conceptualizations created by humans (humans using GenAI), these two researchers' final grades showed correlations of 0.82 (0.76). For the articles on gold futures (commodity CFDs), their correlation was 0.89 (0.89).

To ensure that our four graders were aligned with the criteria at IG, we decided to post-hoc remove any grader whose grades were, on average, 2 units or further from the sample grades provided by the head of web analytics. Fortunately, each of our four graders were within this range.

3.3.3 Surveys. Before and after participating in the experiment, employees completed a survey. In the pre-experimental survey, participants were asked questions related to their demographics (e.g., gender, age, education), personal attitudes (e.g., career commitment), attitudes and behaviors toward GenAI use (e.g., use at work and in personal life, anxiety provoked by GenAI), writing style and use of GenAI in the writing process, job design (e.g., main job tasks, job satisfaction, job design characteristics), and IT skills assessment (e.g., on coding, machine learning, and GenAI theory). A small subset of these questions were asked again at the end of the experiment in the post-survey to capture changes in attitudes. The full survey questions can be found in the supplementary materials.

RESULTS

We now report our analyses on the impact of GenAI on the performance of marketing specialists, technologists, and web analysts in article conceptualization and execution tasks. As summarized in Table 2, our findings reveal varying performance outcomes across experimental conditions, alongside distinct task approaches adopted by participants. The following sections detail two main results. First, we demonstrate that GenAI significantly improves the performance of the marketing and technology specialists for the article conceptualization task and marketing specialists for the article execution task. However, GenAI has no effect on technology specialists for the article execution task. Second, we suggest an explanation to this finding by demonstrating a divergent task resolution approach of technology specialists (with respect to web analysts and marketing specialists) and a different demand of the task of executing the article (rather than conceptualizing it).

	Article Conceptualization	**Article Execution**	**Task Resolution Approach**
Marketing Specialists	GenAI closes the divide with web analysts	GenAI closes the divide with web analysts	Similar to web analysts
Technology Specialists	GenAI closes the divide with web analysts	GenAI does not close the divide with web analysts	Divergent from web analysts

TABLE 2
Comparison of task resolution approaches with and without GenAI, Baseline is web analysts

4.1 GenAI Closes the Divide Across Occupations, Except for Technology Specialists in Article Execution

The experimental results show that GenAI induced substantial improvements in speed and quality across both conceptualization and article creation. For conceptualizations, our findings indicate that without GenAI, neither marketing specialists nor technology specialists could match the quality achieved by web analysts. However, when GenAI was introduced, marketing and technology specialists matched the quality of web analysts' conceptualizations. On the contrary, in article writing, with GenAI support, only marketing specialists produced results that were indistinguishable from web analysts, and both web analysts and marketing specialists wrote articles that were notably better than those of technology specialists. Unfortunately, low compliance does not allow us to make claims about the distance between groups in the control condition for article writing. Appendix A, presents additional analyses performed as robustness checks.

4.1.1 Both Marketing and Technology Specialists closed the divide with web analysts in article conceptualization. GenAI assistance evened performance differences across groups in article conceptualization. Without GenAI, web analysts outperformed both marketing specialists and technology specialists in creating article conceptualizations. However, when assisted by GenAI, marketing specialists and technology specialists produced article conceptualizations on par with web analysts.

The grades for conceptualizations, by condition and job expertise, are depicted in Figure 3. On average, conceptualization grades were 3.13 without, and 4.11 with, GenAI. This difference is significant ($t = 7.53$, $p < .0001$). Further, the significance is maintained when controlling for important demographics (gender, age and education), as well as job characteristics (tenure at IG, GenAI usage) and expertise, as shown in Table B1.

Without GenAI, web analysts created conceptualizations that averaged a grade of 3.82, compared to 3.04 for marketing specialists and 3.02 for technology specialists; these differences were both significant at the 5% level ($t = 3.47$, $p = .001$ and $t = 4.48$, $p = 0.0001$, respectively). However, when equipped with GenAI, web analysts created conceptualizations that averaged a grade of 4.12, compared to 4.18 for marketing specialists and 4.05 for technology specialists; these differences were *not* significant ($t = -0.44$, $p = 0.66$, and $t = 0.48$, $p = 0.63$, respectively). We see these results reflected in a fully interacted regression between experimental condition and expertise in Table B1. Put together, there is evidence that, for task conceptualization, GenAI (1) increases the output quality and (2) closes the quality gap, allowing marketing specialists and technology specialists to match the quality of web analysts.

TABLE 3
Article Conceptualization Outcomes (SEs clustered by employee)

	Quality	Time	Productivity
(Intercept)	3.41***	104.15***	–0.61
	(0.38)	(19.77)	(0.37)
GenAI	1.14***	–49.50***	0.38**
	(0.25)	(6.91)	(0.14)
Task Expert	0.72**	–17.14	0.06
	(0.23)	(11.24)	(0.05)
Novice	–0.12	–9.64	0.10
	(0.24)	(10.81)	(0.07)
GenAI × Task Expert	–0.83**	18.61+	–0.09
	(0.27)	(9.87)	(0.19)
GenAI × Novice	–0.11	14.56	–0.09
	(0.33)	(11.30)	(0.15)
Num.Obs	130	130	129
R^2	0.395	0.324	0.271
Controls	✓	✓	✓

Notes: + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Controls are gender, age, education, tenure, and GenAI usage. Reference group is Domain Expert without GenAI.

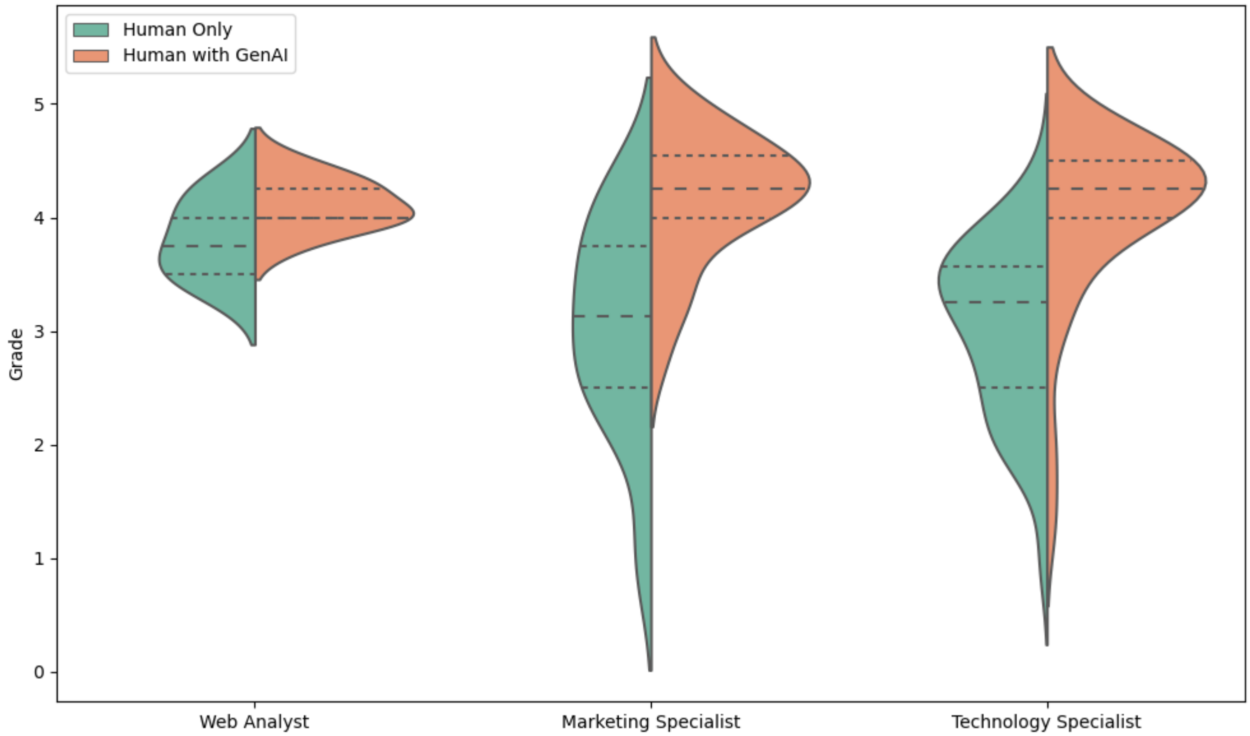

FIGURE 3
Output quality for conceptualizations by condition and expertise. Quality is significantly higher when employees are given access to GenAI; further, GenAI allows marketing and technology specialists to match web analysts. The dashed lines represent the quartiles of the data distribution. Note, due to the small sample size, the data for web analysts does not include all four quartiles.

Our analysis reveals that GenAI significantly narrows performance divides by primarily improving previously lower-performing marketing and technology specialists. This general trend is evidenced by an increase in mean grades and a reduction in the standard deviation (narrowing of the distribution) on article conceptualization under the "Human with GenAI" condition, indicating improved overall performance and consistency (see 3 for visual representation of grade distributions).

Moreover, in the article conceptualization, we were able to conduct a deeper analysis to determine if GenAI helped low-performing individuals match high-performers within each occupation group, rather than across groups. This was possible because the randomization design required each participant to complete an article conceptualization first without GenAI and then with GenAI. We classified employees as low-performers (versus high-performers) if their scores fell below (versus above) the median for their respective expertise group on the conceptualization task without GenAI [7]. We defined low-performers (high-performers) as employees who scored below (above) the median grade for their expertise group on the article conceptualization task without GenAI. For example, the median technology specialist grade on the article conceptualization without AI was 3.25, and 18 technology specialists were designated as high-performers (scored above 3.25) while 14 were designated as low-performers (scored below 3.25) relative to other technology specialists.

Our findings confirm that, without GenAI, significant performance disparities existed within each expertise group for article conceptualization: high-performers achieved significantly higher grades than low-performers for web analysts ($t = 4.56$, $p = .002$) marketing specialists ($t = 6.62$, $p < .0001$) and technology specialists ($t = 8.26$, $p < .0001$). Crucially, when GenAI was introduced, these intra-group performance gaps largely diminished. We found no statistically significant difference in grades between high-performers and low-performers for web analysts ($t = 0.61$, $p = 0.29$), marketing specialists ($t = -0.48$, $p = 0.68$) and technology specialists ($t = -1.1$, $p = 0.85$).

These results demonstrate that, for article conceptualization, GenAI effectively reduces—and in some cases, eliminates—the performance gap between individuals who initially perform well on a task and those who do not within the same expertise group. This phenomenon, particularly the significant improvement of previously lower-performing individuals, is visually depicted in Figure 4, where data points for low-performers (orange dots) are placed more to the top-left with respect to high-performers (blue dots) and above the diagonal line, indicating a higher improvement with GenAI.

We further consider the time required to complete the conceptualization, and find that the positive effects on conceptualization quality did not come at the expense of performance (i.e., productivity was not compromised). We find that employees completed conceptualizations in substantially less time when equipped with GenAI: on average, conceptualizations took 62.8

[7] It is important to note that this specific analysis could not be replicated for the article execution task due to insufficient “Human Only” data for all employees. This analysis was not pre-registered.

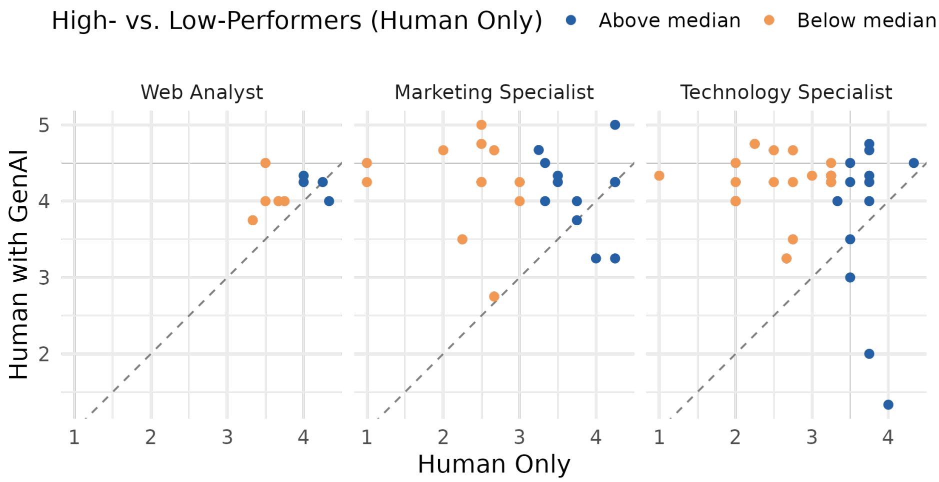

FIGURE 4
Output quality for article conceptualization by condition and expertise. Each employee received two grades for this task: one without AI (x-axis) and one with AI (y-axis). Employees are split by high- and low-performers within group, determined by their grade without AI (above or below the median score within their expertise group).

minutes without, and 23.0 minutes with, GenAI. This difference is significant ($t = 6.87$, $p < .0001$). Further, the significance is maintained when controlling for important demographics (gender, age and education), as well as job characteristics (tenure at IG, GenAI usage) and expertise, as shown in Table B1.

We can further incorporate speed into our analyses by measuring productivity, which we define as the grade a participant earns divided by the time spend completing the task[8]. Across all conditions, GenAI made employees substantially more productive; more details are provided in Table B1. This is not a surprising result: we saw how employees delivered higher quality output with GenAI, and how employees worked much faster with GenAI.

4.1.2 Only marketing specialists closed the divide with web analysts in Article Execution. We find that, once GenAI is introduced, only marketing specialists can produce work on par with web analysts—the distance between web analysts and technology specialists remains statistically significant. In short, GenAI levels the playing field in article execution only for marketing specialists.

[8]We did not pre-register this dependent variable for the simple reason that we did not conceive of it at the time of pre-registration; more details are provided in Appendix B.

The grades for article writing, by condition and job expertise, are depicted in Figure 5. Without GenAI, web analysts wrote articles that averaged a grade of 3.29, compared to 2.81 for marketing specialists and 3.38 for technology specialists. Note that, although the average quality achieved was higher for web analysts than marketing specialists, this difference was not significant at the 5% level ($t = 0.93$, $p = 0.191$). This is a surprising result, which we believe is driven by a small sample size of web analysts in the experimental group. After removing non-compliers, we retain only four articles created by two web analysts without GenAI; therefore, any comparison tests involving web analysts are severely underpowered.

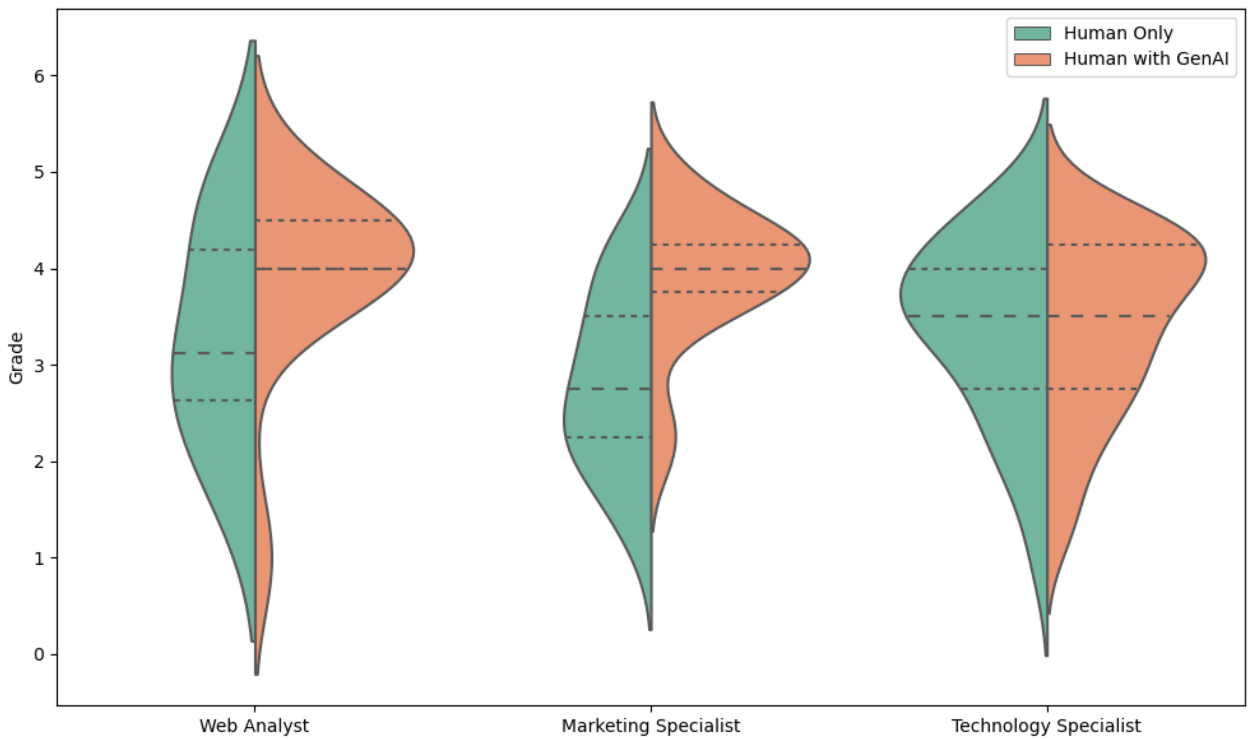

FIGURE 5
Output quality for article writing by condition and expertise. Quality is significantly higher when employees are given access to GenAI; further, GenAI allows marketing and technology specialists to match web analysts. The dashed lines represent the quartiles of the data distribution. Note, due to the small sample size, the data for web analysts does not include all four quartiles.

When equipped with GenAI, web analysts wrote articles that averaged a grade of 3.96, compared to 3.92 for marketing specialists and 3.42 for technology specialists; the difference

TABLE 4
Article Outcomes (SEs clustered by employee)

	Quality	Time	Productivity
(Intercept)	2.89***	141.73*	–0.66
	(0.64)	(55.49)	(0.45)
GenAI	1.29***	–80.24+	0.26
	(0.36)	(41.39)	(0.16)
Task Expert	0.58	18.13	–0.26*
	(0.70)	(41.65)	(0.13)
Novice	0.67+	–41.53	0.09
	(0.38)	(43.19)	(0.10)
GenAI × Task Expert	–0.63	12.26	0.24
	(0.77)	(51.97)	(0.24)
GenAI × Novice	–1.26**	29.64	0.31
	(0.43)	(44.63)	(0.20)
Num.Obs	144	144	142
R^2	0.203	0.357	0.267
Controls	✓	✓	✓

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Controls are gender, age, education, tenure, GenAI usage, and whether the article was written using a brief created with the help of AI. Reference group is Domain Expert without GenAI. One expert responded “many" when asked for their numeric tenure; despite this unsatisfactory data, we decided not to remove them because of our already low sample size. Instead, we assigned them the average tenure of experts in our sample.

between web analysts and marketing specialists was not significant (t = 0.10, p = 0.459), while the difference between web analysts and technology specialists was nearly significant at the 10% level (t = 1.72, p = 0.11). Once again, we believe the lack of 5% significance between web analysts and technology specialists is due to a small sample size retained in the web analysts group; the difference between marketing specialists, who scored similarly to web analysts, and technology specialists, was significant (t = 2.76, p = .007).

We see these results reflected once more in the fully interacted regression in Table 4. Put together, there is suggestive evidence that, for article writing, GenAI (1) increases the output quality for web analysts and marketing specialists and (2) shrinks the quality gap between web analysts and marketing specialists, but not between web analysts and technology specialists.

We then focused on the time required to complete an article to ensure that the effect on quality did not come at the expense of performance (i.e., that productivity was not compromised). We find that employees equipped with GenAI wrote articles much faster than employees not equipped with GenAI. On average, writing an article took 87.2 minutes without, and 22.4 minutes with, GenAI. Once again, this difference is significant (t = 5.22, p < .0001), and the significance is maintained when controlling for important demographics (gender, age and education), as well as job characteristics (tenure at IG, GenAI usage), as shown in Table 4.

Again, our analysis reveals that GenAI significantly narrows performance divides by primarily boosting previously lower-performing marketing specialists. This general trend is evidenced by an increase in mean grades and a reduction in the standard deviation (narrowing of the distribution) on article execution tasks under the "Human with GenAI" condition, indicating improved overall performance and consistency. However, contrary to article conceptualization, this pattern is particularly evident only for web analysts and marketing specialists. However, a key exception is observed for technology specialists, where GenAI—despite leading to a slight improvement in their mean grade—did not effectively reduce performance variability or fully close their performance gap, meaning the distribution of high-low performers remained wide, unlike in all other observed conditions (5).

4.2 Why Only Low-Performing Marketing Specialists Were Able to Use GenAI to Close the Divide in Article Writing

Our experimental results reveal a core puzzle: GenAI helped low-performing marketing specialists, but not low-performing technology specialists, close the gap with web analysts in executing articles. Because individual experimental outcomes were not linked to interview transcripts to preserve anonymity, we can suggest an explanation for this dynamic by contrasting the across-group differences of marketing specialists, technology specialists, and web analysts.

4.2.1 Different knowledge distance from web analysts relates with a different task resolution approach. One key finding from our interview analysis is that knowledge distance shaped a different task resolution approach: for their closer knowledge distance, marketing specialists' methods more closely resembled the approach used by web analysts, whereas technology specialists diverged more substantially from both groups. In this section, we review how the knowledge distance influenced the approaches to resolve tasks.

Web analysts' task resolution approach was centered around the use of SEO (Search Engine Optimization) and CRO (Conversion Rate Optimization) principles to structure and write web articles. SEO principles were guiding them on ensuring articles ranked highly in search engine results—particularly on Google—thereby reaching a broader online audience. One web analyst highlighted the importance of optimizing content to align with Google's algorithm:

> "In this job, you learn everything about SEO. It's all because Google essentially ranks your content. If you don't optimize for that, you won't get traffic. Without traffic, you won't get customers or revenue. Naturally, Google is an algorithm, not a human. By writing in a certain way—applying keywords and signaling to Google that your content is high-quality. That helps you rank better, drive more traffic, and gain greater brand visibility for our stuff." (Web Analyst - TE_04)

Another web analyst explained us the importance of writing articles not just for readers but also for the search engine algorithm, emphasizing that it is not only important what keywords you identify in the conceptualization, but also how those keywords are strategically include in specific parts of the copy—particularly in headings:

> “To write a good article, you need to focus a lot on headings—like H1, H2, and so on — which are key elements that Google evaluates. They also have a solid grasp of keywords, which are critical in SEO. Doing it regularly helps you understand what makes an article effective or not.” (Web Analyst - TE_02)

Beyond SEO, web analysts also pointed to the importance of CRO. While SEO helps attract readers, CRO aims to convert them into active users by inviting them to complete a specific action, such as making a purchase or signing up for a newsletter. A web analyst explained:

> “To write a good article, I believe you need some creativity in understanding the customer flow—how customers think and what their intent is. [. . .] you might craft your article as a bridge between two different moments on the website. A customer may arrive from another page, read the article, and become convinced to act. At that point, you must keep the flow going—adding a call to action at the bottom to lead them toward what you’ve convinced them to do. Conversion is crucial, and the call to action is vital because it ensures the flow continues after the reader finishes.” (Web Analyst - TE_04)

In sum, web analysts had a clear language and technical approach to task resolution, reflecting their hands-on, repeated experience on the task.

Marketing specialists, when describing their approach to the task, focused on broader marketing considerations, such as customer personas and overall campaign objectives. One marketing specialist illustrated this perspective:

> “Writing an article is essentially a content writing task for customers. No matter what area you’re in—events, communications, or CRM—you still need to know how to write content for customers. It’s the same language. You just have to add certain elements to rank higher on Google.” (Marketing Specialist - DE_04)

Another marketing specialist said:

> “When I structure content, I try to put myself in the customer’s shoes. I consider their journey to determine what information they need to understand the content or requirements and take action.” (Marketing Specialist - DE_03)

Whereas web analysts devoted significant effort to optimizing content for search engines and conversions, Marketing Specialists took a broader marketing perspective to engage audiences.

Like web analysts, marketing specialists emphasized keeping the customer at the center of their writing choices. They hence had a similar foundational knowledge to web analysts. However, unlike web analysts, marketing specialists did not explicitly mention SEO or CRO technicalities—proving that they were missing the hands—on experience that web analysts accumulated in producing web articles daily in their jobs.

Technology specialists took a fundamentally different approach compared to web analysts and marketing specialists. While web analysts and marketing specialists often focused on the customer's perspective in their work, technologists did not mention this during their interviews. Instead, many drew parallels between crafting web articles and writing technical documentation, such as internal wiki pages—an approach emphasizing the value of quick, straightforward sentences to convey the article core message. Intrinsically, they claimed they were transferring their technical writing style in technical documentation to web-article writing. For example, one data scientist summarized what a good technical document would contain:

> "I never wrote web articles during my two years here, but we do produce a lot of technical documentation, especially where explaining our math models is crucial. Over time, I've become quite skilled at summarizing these technical documents in a non-technical way so that higher-up managers, who might not fully understand what we do, can still grasp the content." (Data Scientists - NO_01)

The underlying approach to task resolution of technology specialists reveals a lack of solid hands-on experience (not aligning with SEO, CRO) and foundational knowledge (they missed the the marketing specialists' sensitivity of what constitutes a strong marketing artifact) compared to the other two occupations.

4.2.2 Difference in article conceptualization and article writing shifts the problem nature from structured abstraction to ill-structured embodiment. A second finding of interview analysis is that participants drew a clear line between the characteristics of article conceptualization and execution. They defined conceptualization as the "abstraction" of an article's features—a highly structured process tightly guided by a codified template. In contrast, they saw execution as the “embodiment” of ideas into tangible prose—an inherently unstructured task, precisely because it lacked a formal template to shape it.

Web analysts emphasized that article conceptualization required to follow a strict “template” used to produce the output as the main feature of the article conceptualization task:

> “The article conceptualization is basically a template you fill in. It’s similar to what you do at school when you have fields for name, surname, project title, and so on. It’s a template that specifies exactly what the article should cover.” (Web Analyst - TE_05)

In stark contrast, web analysts framed the act of writing as one of embodiment, requiring a feel for the craft to give the abstract facts a physical form:

> “There’s freedom for creativity in the article itself. You basically need to turn the conceptualization into sentences and paragraphs. Before GenAI, you really had to start from a white page.” (Web Analyst - TE_04)

The same task distinction was also reported by the marketing specialists who took part to the experiment. One marketing specialist explicitly contrasted the simplicity of the brief compared to writing the article itself:

> “Writing a schematic brief was much simpler than writing the article. When I had to draft the conceptualization, I essentially outlined instructions, whereas writing the article demanded me to do an extra step and ensure the content was logical, engaging, and coherent.” (Marketing Specialist - DE_03)

This distinction between article conceptualization and writing lies indeed in the difference between listing the features of the article and producing and article (paragraphs and sentences) which contain those features. One data scientist noted:

> Briefs had headings—Heading 1 and Heading 2—and a particular template to follow. If it says, ‘I need a heading, a URL, and a subheading,’ then the instructions are more defined, and I can understand what to do. Articles were a blank page. The more open-ended something is, the more room there is for error. (Data Scientist - NO_03)

In sum, our interviews with web analysts revealed that conceptualization and execution demand fundamentally different abilities. We characterize conceptualization as an act of structured abstract formulation—a process of listing the article features following the template required. In contrast, execution is an act of embodiment: the craft of writing the paragraphs and sentences that give a tangible shape to the abstract concept.

4.2.3 How Task Nature Interacted with the Task Resolution Approach. Finally, our results show that both marketing specialists and technology specialists found article writing much more complex that article conceptualization. As one marketing specialist observed:

> "Since I'm not constantly engaged in the same tasks, it sometimes takes me longer to grasp specific details—like selecting in-depth keywords for conceptualizations—compared with someone who works on that every day. I've noticed a significant difference when I work independently versus when I use GenAI assistance." (Marketing Specialist - DE_03)

This difficulty in the process was, in fact, the result of marketers lacking the practical hands-on experience needed to promptly devise a solution for the task at hand. Similarly, the ill-structuredness of the second task manifested as a practical challenge for the technology specialists. One data scientist noted:

> The tasks varied greatly from one another. Briefs were simple—I just followed the steps outlined in the document. They were very similar to what we do: clear and concise. The problem was the writing itself: I did not know where to start! (Software Developer - NO_05)

Another technology specialist expressed frustration with the very nature of the craft:

> "I think it was clear from the template what I had to write. Yet, I struggled to weave them naturally into the narrative. I knew I had to include buzzwords and catchy phrases, but incorporating them seamlessly into the storytelling was quite complicated and the sentences I produced looked unreadable."(Data Scientists - NO_03)

Yet, even though both groups found article writing difficult, their different approaches led them to relate to GenAI in fundamentally different ways. Marketing specialists, for instance, struggled to come up with sentences and paragraphs because they lacked hands-on experience. However, since they possessed the foundational knowledge to distinguish good copy from bad, they used GenAI to accelerate idea generation:

> "Without GenAI, things would have been much worse for me. I had an idea of what the article should look like. I can tell you if one article is better than another. But writing it? That's a different story if you don't write regularly. So, yes, doing without GenAI would have been painful." (Marketing Specialist - DE_04)

In contrast, technology specialists—lacking both hands—on experience and foundational knowledge—often edited GenAI outputs in ways that conflicted with the rubric established by the head web analyst. Whereas marketing specialists used GenAI primarily to generate *ideas*, technologists relied on it to produce *content*, which they then revised by fixing grammar and syntax or shortening sentences for clarity—often cutting material they did not recognize as relevant. One data scientist, for example, praised GenAI for helping him incorporate a "marketing spin" into his articles:

> **Data Scientist -** NO_01: "I think GenAI helped me adding that "marketing spin" to my work."
>
> **Interviewer**: "What do you mean by 'Marketing spin'?"
>
> **Data Scientist -** NO_01: "GenAI suggested some catchy hooks for the start of my articles. For example, when writing about trading gold futures, GenAI proposed catchier language and seemed a nice way into the article. . . [pause], actually? I didn't fully understand what it was doing because I never wrote an article like that. I added random stuff to make it more 'marketing', you know what I mean?" (Interview with Data Scientist - NO_01)

As another example, technologists emphasized that good web articles should be concise, taking issue with the GenAI output for being overly verbose and deciding to shorten it. As one data scientist explained, he removed large portions of the AI-generated content, judging them unnecessary for enhancing the article's value:

> "I felt GenAI put so many words into the diluted the technical content. I ended up removing those parts to keep the article straightforward and to the point." (Interview with Data Scientists - NO_04)

As such, some of them even expressed annoyance at the "fancy language" generated by the GenAI tool:

> "What I can say is that Gen AI didn't really improve the quality of my articles. I ended up reversing most of its suggestions because I prefer articles that are clear and direct. As I mentioned, whenever Gen AI offered something that felt too buzzword-heavy or non-technical, I simply removed it." (Data Scientists - NO_02)

To probe him further into this example, we asked him to clarify what he meant. In his reply, he cited a final call to action inserted by GenAI at the end of the article — a sentence he himself did not precisely recall. When asked whether a call to action should be part of the article, as CRO principles would suggest, he maintained that the article should simply describe the topic and convey the message. In his view, the assignment was strictly to write an article on the given subject, not to prompt a specific action from readers.

Approaching the articles as if they were wiki pages, according to our evaluation rubric, may have diminished the overall quality of the article. For example, the head of web analytics emphasized that web articles should be roughly 800 words to meet SEO and CRO standards—an approach also built into IG's bespoke GenAI tool, which was designed to produce text of this length. Although this approach was highly effective for technical documentation—ensuring precise and unambiguous explanations did not align with the approach of web analysts, which centered on attracting and engaging an external audience through marketing-oriented considerations like CRO and SEO. Developers, however, proved to be unaware of what a good article was.

In sum, the analysis of the interview transcripts provided a potential explanation for why only marketing specialists were able to catch up with web analysts by using GenAI. Technology specialists seemed to be unable to leverage GenAI to close the performance gap with web analysts, unlike marketing specialists, because they lacked the foundational knowledge that enabled effective use of GenAI's recommendations. While marketing specialists shared core marketing principles with web analysts, technology specialists relied on a different evaluative framework rooted in writing technical documentation. This led them to prioritize concise, straightforward writing in web articles—often at odds with the SEO and conversion rate optimization (CRO) strategies critical for web content success.

DISCUSSIONS

The central result of our study is that GenAI may better equip outsiders to redeploy their knowledge and close the divide with insiders (as predicted in our 1). At IG, there were adjacent outsiders (e.g., marketing coordinators, CRM specialists) and distant outsiders (e.g., data scientists or software developers) who, without GenAI, were not able to match the quality of insiders (i.e., web analysts). We show that GenAI removed those "low-performing" outsiders, with one exception: distant outsiders on the task that had an unstructured (rather than structured) process and required the embodiment of a concept into a tangible artifact (rather than abstract conceptualization).

These findings contribute to ongoing debates about the effect of GenAI on work, which remains polarized around conflicting sides. One side of the debate contends that GenAI will undercut the value of human expertise (e.g., Brynjolfsson et al., 2024; Dell'Acqua et al., 2023). Another side suggests that GenAI will not produce meaningful changes in work and organizations (e.g., Otis et al., 2025), pointing out, for example, that the divide between high- and low-performing experts hinges on certain abilities that GenAI will never be able to master.

We contribute to this discussion by exploring the "wall effect" of GenAI—that is, the point at which GenAI can no longer meaningfully reduce the expertise gaps between insiders and distant outsiders (as opposed to adjacent outsiders). Our findings indicate that while GenAI does help adjacent outsiders (e.g., marketing specialists in our empirical context) perform tasks at a level typically performed by insiders (e.g., in our empirical context, web analysts), it nonetheless hits a wall in bridging the divide between the performance of distant outsiders (in our empirical context, technology specialists) and insiders. Specifically, we find that the "GenAI wall" emerges from the interaction between the knowledge distance separating different occupations and the demands of various tasks: GenAI is more likely to hit a threshold—or a wall—for distant rather than adjacent outsiders, and for execution rather than conceptualization tasks. We conceptualize this threshold as the GenAI Wall: the point at which AI support is robust enough to enable adjacent outsiders, but not distant outsiders, to surmount tasks that historically required specialized human expertise. In other words, the GenAI Wall Effect delineates the zone where AI can effectively bridge expertise gaps from the zone where deep domain knowledge remains indispensable.

By casting these findings in terms of learning transfer, we advance the distinction between horizontal and vertical transfer of expertise in an organizational context. Vertical transfer denotes the flow of expertise between different levels of mastery within the same domain (for example, from a senior expert to a junior novice), whereas horizontal transfer refers to expertise flowing across occupational communities. In short, horizontal transfer involves applying knowledge across different but comparable occupations and contexts. Classical learning theory suggests that transferring expertise is easier within a domain than across domains, because shared context and schemas facilitate "near transfer." In contrast, cross-domain or "far" transfer is notoriously difficult since knowledge is often situated in specific practices (Lave and Wenger, 1991). Our findings reveal that GenAI can serve as a new conduit for horizontal transfer of expertise that traditionally would not readily cross occupational boundaries. Our findings suggest that adjacent outsiders—individuals from a different but related occupation—were able to perform on par with insiders in conceptualization tasks when aided by GenAI. This implies that GenAI systems can encapsulate and deliver explicit knowledge from one domain to a non-expert in another domain, effectively allowing outsiders to access and apply domain insights horizontally. In essence, GenAI acted as a knowledge intermediary, translating and supplying expertise across what would normally be rigid occupational lines. This horizontal knowledge transfer via GenAI is a novel phenomenon: traditionally, an outsider would have needed extensive training or collaboration with an insider to contribute meaningfully to perform tasks in a different occupational domain. GenAI assistance dramatically shortens that process by providing on-demand, context-specific information and suggestions, thereby democratizing the conceptual phase of problem-solving across roles. These findings extend classic learning transfer theory into the realm of human-AI collaboration: unlike traditional training, where a person gradually internalizes skills, here the GenAI system instantly provides the performance benefits of those skills.

Our study also uncovers the boundary conditions of this expertise—leveling effect showing that GenAI does not uniformly benefit all users or all tasks. For relatively well-structured problems that resemble the data on which the AI was trained, less-skilled individuals can leverage AI suggestions to achieve performance closer to expert levels. This reflects a form of *vertical* transfer of expertise: the AI provides novices with the building-block knowledge or solutions, enabling them to perform at a higher level than their personal experience would otherwise allow. But our findings also suggest a *horizontal* transfer of expertise between occupational insiders and adjacent outsiders, and limits to such horizontal transfer of expertise between occupational insiders and distant outsiders. Thus, when users attempt to apply GenAI beyond the current "wall" of its capabilities, the expertise divide may persist or widen. This nuance enriches the theoretical understanding of when AI augments human capabilities versus when it falls short, emphasizing that the transformative potential of AI is bounded by the nature of the task and the context of its use.

These boundary conditions on the nature of tasks further situate our contribution within the literature examining how GenAI's effectiveness varies across work (Dell'Acqua et al., 2023; Eloundou et al., 2024). This research highlights the "jagged frontier" between tasks that AI can and cannot perform. We extend this discussion by showing that, for tasks situated within this frontier, the effectiveness of the human–AI ensemble depends on the user's knowledge distance from the task. This factor becomes particularly consequential as work progresses from conceptualization to execution, where domain-specific expertise interacts more directly with AI-generated outputs (Amabile et al., 1988; Mannucci and Perry-Smith, 2022). Notably, these results allow us to reconceptualize the recently identified "ideation–execution gap" (Si et al., 2025). This prior research found that AI-generated ideas, while initially novel, often diminish in value once experts attempt to implement them, suggesting the ideas themselves lack feasibility. Our study challenges this explanation by shifting the causal focus from the idea to the implementer. We argue the gap emerges not simply because an AI's idea fails the test of reality, but because execution is a complex task whose difficulty is magnified by a user's knowledge distance. Our findings substantiate this: the performance gap between outsiders and insiders was closed in the ideation phased but not during execution—and critically, only for distant outsiders. The success of adjacent outsiders proves the bottleneck was not the inherent quality of the idea, but the lack of deep, procedural knowledge needed to execute it effectively.

Our study also contributes to a deeper understanding of the evolving boundary conditions of occupational expertise in the age of AI. Traditionally, occupations maintain distinct jurisdictions defined by specialized knowledge and skill. Those boundaries are reinforced by the fact that without lengthy training, an outsider cannot perform an insider's tasks effectively. GenAI upends this logic by externalizing and disseminating expertise. Our findings illustrate that what counts as "expertise" in a role is increasingly split between human judgment and AI-provided knowledge. The exclusive domain of an occupation shrinks for tasks that AI can handle, as non-experts equipped with a versatile AI tool can achieve competent results in those areas. Conversely, the areas of expertise that remain exclusive—the tasks on the far side of the GenAI Wall—become even more critical to what defines an expert. For example, if drafting a basic marketing plan or writing code can be done by anyone with AI assistance, the unique value of a professional marketer or programmer may shift to higher-order skills like strategic decision-making, novel innovation, or complex problem-solving that AI cannot yet replicate. In theoretical terms, occupational expertise is becoming more conditional: its importance is preserved primarily at the frontier where AI support stops. Our work thus refines the boundary conditions of expertise by identifying the role of technology in mediating those boundaries. Rather than viewing technology simply as a tool that experts use, we portray it as an active agent that reconfigures the distribution of expertise across roles. This perspective extends research on occupations and technology by demonstrating that GenAI can serve as a partial substitute for formal expertise in some contexts, thereby triggering a need to redraw professional boundaries and reevaluate how expertise is credentialed and valued in organizations. We expect that occupations will respond to this pressure by renegotiating jurisdictions—perhaps focusing on the emergent tasks that AI cannot do—and by developing new norms around the acceptable use of AI in professional work. In short, the GenAI Wall Effect advances understanding of occupational boundaries by pinpointing where technology blurs them and where it reinforces the remaining dividing lines of expert authority.

Our study thus identifies important boundary conditions for when GenAI can mitigate occupational expertise gaps. We show that the effectiveness of GenAI's knowledge mediation is contingent on task type and the knowledge distance between actors. Not all tasks benefit equally from AI support, and not all knowledge gaps can be bridged. Our study's findings suggest that the ability of GenAI to reduce performance disparities was evident in conceptualization tasks (e.g. idea generation, problem framing) but not in execution tasks (e.g., detailed implementation, hands-on problem-solving). This task dependence reveals a boundary condition rooted in the nature of knowledge required: conceptual tasks primarily demand explicit knowledge, general reasoning, and creative recombination of ideas—precisely the kind of content that GenAI and LLMs can provide. Execution tasks, by contrast, draw much more on tacit knowledge, context-specific judgment, and practical skill that are acquired through experience (Lave and Wenger, 1991). From a transfer-of-learning perspective, GenAI can supply outsiders with information and heuristics that enable far transfer for abstract thinking, but it cannot easily impart the embodied competences needed for near transfer in applied work (Singley and Anderson, 1989). In essence, when a task's knowledge requirements are more abstract and codifiable, AI assistance goes a long way; when tasks involve concrete application and context-bound nuances, the outsider remains at a disadvantage because they lack the lived experience to interpret and enact the AI's advice. This finding echoes and enriches situated learning theory: even with an advanced AI tutor, true mastery of execution may require legitimate peripheral participation and time spent in a "community of practice" to develop intuition and skill. GenAI can provide the map, but navigating the terrain is another matter.

Another boundary condition we highlight is the degree of occupational knowledge distance between the person and the task domain. Our study deliberately examined "adjacent outsiders"—individuals whose own field is related enough to have some shared language or analytical frameworks with the target domain. The fact that GenAI helped these adjacent outsiders in ideation suggests that a moderate knowledge distance can be bridged by AI-provided knowledge. However, even for these relatively close outsiders, AI assistance did not erase the performance gap in execution tasks. This implies that as the knowledge distance grows (from insider to adjacent outsider, to perhaps a far outsider with no relevant background), GenAI's utility likely diminishes further, especially for complex tasks. An outsider who cannot understand or vet the AI's suggestions will struggle to implement them effectively. Thus, a theoretical boundary condition emerges: GenAI is most transformative as a support tool when the user has at least a minimal foundation in or affinity with the target knowledge domain. We contribute here by suggesting where GenAI's "augmented expertise" ends and where true domain expertise remains indispensable. In practical terms, organizations cannot assume that GenAI will allow any employee to perform any task with equal proficiency; rather, the technology's benefits accrue under specific conditions—when tasks are of a type that AI can handle (conceptual, knowledge-intensive, decomposable into text or rules) and when users are within a reasonable knowledge distance from understanding the focal task domain.

By detailing these boundary conditions, we refine emerging theories of AI in organizations, cautioning against one-size-fits-all claims. GenAI can indeed challenge traditional expertise distributions, but only in certain arenas and under certain boundary conditions. This insight contributes to the literature on tasks, occupations, and expertise by illustrating how new technology can blur expertise boundaries in some contexts while reinforcing them in others (Abbott, 1988). It also adds nuance to the literature on human–AI collaboration, showing that the impact of AI on human performance is highly context-dependent, moderated by the nature of tasks and the user's prior knowledge.

BROADER IMPLICATIONS, LIMITATIONS, AND CONCLUSIONS

Our findings on GenAI's capacity for horizontal expertise transfer raise profound questions for the foundational principles of organizational design. The fact that GenAI enables adjacent outsiders to perform insider tasks—contingent on knowledge distance and task type—suggests a potential disruption to how we theorize expertise and structure work. We elaborate on two key implications for future research, moving from how GenAI transforms work to how it may reshape organizations.

First, our results suggest that the very principles of specialization may be changing. The traditional path to expertise relies on repeated, hands-on practice within a specific role to build tacit knowledge (DiBenigno, 2018; Ericsson and Charness, 1994). We challenge this by showing that GenAI may allow individuals to bypass this learning curve. Our findings imply a shift where the critical skill is no longer just task-specific experience, but a broader domain knowledge that allows an individual to effectively direct, interpret, and refine AI outputs. This "expertise transfer" effect—from insiders to adjacent outsiders—suggests that organizations may need to redefine what constitutes an "expert," potentially prioritizing flexible, domain-level understanding over narrow, task-level repetition and rewarding individuals accordingly.

This evolving nature of expertise, in turn, sparks new research questions at the crossroads of GenAI and the division of labor. For centuries, the logic of organizational structure has been to cluster tasks into stable jobs to overcome the learning curves associated with specialization (Smith, 1776; Simon, 1945; Hayek, 2013). This division of labor is a solution to the high cost of transferring deep knowledge. Our findings, however, suggest GenAI may fundamentally alter this logic. By flattening the learning curve for adjacent outsiders, GenAI could enable a more fluid and dynamic allocation of tasks, increasing intra-organizational mobility without the productivity losses typically associated with reassigning workers to unfamiliar roles (Cohen and Bui, 2024; Postrel, 2002). If an employee with strong foundational knowledge can perform tasks in an adjacent domain with AI assistance, the rigid boundaries between jobs may become less efficient than a more flexible, skills-based orchestration of work.

Yet, we must acknowledge the limitations of our study, which themselves point to critical avenues for future research. Our analysis is confined to a single task and a single direction of cross-occupational transfer. It is entirely possible that a GenAI wall would emerge under different conditions—for example, if web analysts were asked to perform technologists' tasks. Furthermore, our study examines peer-to-peer and not manager-subordinate expertise transfer: we do not investigate whether GenAI enables subordinates to take on managerial tasks (or vice-versa, as in Barley (1996)), leaving open vital questions about how GenAI might reconfigure hierarchical structures and career progression. Future field research is therefore essential to map the true contours of a GenAI wall across different tasks, roles, and organizational levels to understand when GenAI acts as a bridge for talent and when it reinforces the need for deep, incumbent expertise.

This experiment suffered also from non-compliance with the intention to treat. We can however claim that the experimental balance remains unaffected, and we discuss at length in Appendix A why these limitations do not pose a threat to our theorizing. We also note deviations from the pre-registration, as fewer employees than expected participated in the experiment. Nevertheless, while our sample size is limited, all of our statistical analyses fully account for the size of the study.

Finally, there is the question of whether firms will actually reorganize following the evidence that GenAI modifies individual knowledge walls. From our lab-in-the-field study, one might infer that, since the expertise required to perform organizational task is much more foundational and less task-specific, we could expect increased occupational mobility and battles between occupations to affirms new jurisdictions. Yet, there could be elements overlooked by this paper that ultimately will distinguish between what firms will actually do from what our experiment prove they may be able to achieve with GenAI.

REFERENCES

Abbott, A. 1988. The system of professions: An essay on the division of expert labor.

Agrawal, A., Gans, J., and Goldfarb, A. 2018. ***Prediction Machines: The Simple Economics of Artificial Intelligence***. Harvard Business Review Press.

Amabile, T. M. et al. 1988. A model of creativity and innovation in organizations. ***Research in organizational behavior***, 10(1): 123–167.

Autor, D. 2024. How ai could help rebuild the middle class.

Autor, D. H. and Handel, M. J. 2013. Putting tasks to the test: Human capital, job tasks, and wages. ***Journal of labor Economics***, 31(S1): S59–S96.

Barley, S. R. 1996. Technicians in the workplace: Ethnographic evidence for bringing work into organizational studies. ***Administrative science quarterly***, 404–441.

Berger, P. G., Cai, W., Qiu, L., and Shen, C. X. 2024. Employer and employee responses to generative ai: Early evidence. ***Available at SSRN 4874061***.

Brynjolfsson, E., Li, D., and Raymond, L. R. 2024. Generative ai at work. ***Quarterly Journal of Economics***.

Cohen, L. E. and Bui, L. H. 2024. The coevolution of tasks and expertise.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. 2023. Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. ***Harvard Business School Technology & Operations Mgt. Unit Working Paper***, (24-013).

DiBenigno, J. 2018. Anchored personalization in managing goal conflict between professional groups: The case of us army mental health care. ***Administrative Science Quarterly***, 63(3): 526–569.

Doshi, A. R. and Hauser, O. P. 2023. Generative artificial intelligence enhances creativity but reduces the diversity of novel content. ***arXiv preprint arXiv:2312.00506***.

Dreyfus, H. L. and Dreyfus, S. E. 2005. Peripheral vision: Expertise in real world contexts. ***Organization studies***, 26(5): 779–792.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. 2024. Gpts are gpts: Labor market impact potential of llms. ***Science***, 384(6702): 1306–1308.

Ericsson, K. A. and Charness, N. 1994. Expert performance: Its structure and acquisition. ***American psychologist***, 49(8): 725.

Garicano, L. 2000. Hierarchies and the organization of knowledge in production. ***Journal of political economy***, 108(5): 874–904.

Hayek, F. A. 2013. The use of knowledge in society. In ***Modern understandings of liberty and property***, 27–38. Routledge.

Hollister, M., Karunakaran, A., and Cohen, L. E. 2023. An ecological model of task disruption: The impact of partial automation of jobs through ai. In ***Academy of Management Proceedings***, volume 2023, 13876. Academy of Management Briarcliff Manor, NY 10510.

Jia, N., Luo, X., Fang, Z., and Liao, C. 2024. When and how artificial intelligence augments employee creativity. ***Academy of Management Journal***, 67(1): 5–32.

Lave, J. and Wenger, E. 1991. ***Situated learning: Legitimate peripheral participation***. Cambridge university press.

Mannucci, P. V. and Perry-Smith, J. E. 2022. "who are you going to call?" network activation in creative idea generation and elaboration. ***Academy of Management Journal***, 65(4): 1192–1217.

Noy, S. and Zhang, W. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. ***Science***, 381(6654): 187–192.

Otis, N., Clarke, R., Delecourt, S., Holtz, D., and Koning, R. 2024. The uneven impact of generative ai on entrepreneurial performance.

Otis, N. G., Haas, B., Clarke, R., Business, H., Soì, S., Delecourt, E., Holtz, D., and Koning, R. 2025. The uneven impact of generative ai on entrepreneurial performance.

Postrel, S. 2002. Islands of shared knowledge: Specialization and mutual understanding in problem-solving teams. ***Organization science***, 13(3): 303–320.

Rousseau, D. M. and Stouten, J. 2025. Experts and expertise in organizations: An integrative review on individual expertise. ***Annual Review of Organizational Psychology and Organizational Behavior***, 12(2025): 159–184.

Si, C., Hashimoto, T., and Yang, D. 2025. The ideation-execution gap: Execution outcomes of llm-generated versus human research ideas. ***arXiv preprint arXiv:2506.20803***.

Simon, H. A. 1945. ***Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization***. New York: Macmillan.

Singley, M. K. and Anderson, J. R. 1989. ***The transfer of cognitive skill***. Number 9. Harvard University Press.

Smith, A. 1776. ***An Inquiry into the Nature and Causes of the Wealth of Nations***. London: W. Strahan and T. Cadell.

Yue, D., Hamilton, P., and Bojinov, I. 2022. ***Nailing Prediction: Experimental Evidence on the Value of Tools in Predictive Model Development***. SSRN.

APPENDIX A: ROBUSTNESS CHECKS AND ADDITIONAL ANALYSES ON THE EXPERIMENTAL RESULTS

We discuss several additional analyses to ensure the robustness of our results. First, we discuss how we adjusted for the non-compliance of certain participants in Appendix A. Second, we perform robustness checks on the quality analysis to ensure that the authors' grades did not influence the results. Even though all of the briefs and articles were anonymized prior to grading, the authors were familiar with the experiment design. We conducted the same analyses as presented in the main using grades only from our MBA graders who were not aware of the nature of the experiment — including that some briefs and articles were created with the assistance of GenAI. More discussion is available in Appendix A.

Non-Compliance

Analysis Using Only MBA Graders

As detailed in Section 3.3, we enlisted four graders to measure the quality of employee output: two authors of this paper and two MBA graders. All briefs and articles were fully anonymized during grading; however, the two authors naturally knew the nature of the study, including the fact that some employees completed their tasks with the help of GenAI.

The MBA graders, by contrast, had no information about the study whatsoever. They were not told that some briefs and articles were completed only by humans, and some with the assistance of GenAI. Fortunately, if we only use the grades from the two MBA graders, our overall results still hold. The average conceptualization grades without GenAI were 2.97, 3.0 and 3.5 across technology specialists, marketing specialists, and web analysts, respectively, and the average grades with GenAI were 4.19, 4.13 and 3.94 with GenAI, respectively. That is, once again, web analysts had the highest average grade on briefs without GenAI, but GenAI closed the quality gap.

For articles, the average grade without GenAI was 3.42, 3.10 and 3.42 for technology specialists, marketing specialists, and web analysts, respectively, and the average grades with

GenAI were 3.56, 3.80 and 3.96, respectively. Once again, performance was similar without GenAI across different expertise groups. The same causes discussed in Section 4.1.2 apply here.

Reconstructed GenAI Output

IG developed two distinct bespoke GenAI tools for creating article briefs and web articles respectively. Employees interacted with the tools by entering inputs described in detail in Section 3.2, and receiving output in the form of generated brief and articles. Participants could download the output briefs and articles as a Word document to make necessary edits and modifications.

While IG did not store the finalized output document, they stored the logs that captured some of the interactions of the tools with the respective LLM API (e.g., OpenAI). From these logs, we were able to partially reconstruct the briefs and articles. While the logs did not store all the information needed to fully construct the original output document, we were able to reconstruct briefs and articles that approximated that original output.

Note, not all employee interactions with the tools were captured in the logs. While some participants were assigned to use GenAI, there are no logs tracing their interactions. However, we still have reason to believe that they used the tool as instructed; for example, we have survey responses that indicate an employee used the tool despite no logs to confirm that is the case. IG confirmed that some employee logs were missing from the dataset. Therefore, the conclusions that can be drawn from the briefs are limited.

The reconstructed briefs and articles were evaluated for quality by a research associate using the same rubric as the completed briefs and articles. Similar to the MBA graders, the research associate had no knowledge of the experimental design prior to grading. Many of the employees had logs that indicated they created multiple versions of the brief or article for the same topic. While it is possible that they only used one of the output documents as a basis for their finalized task submission, it is also possible that they selectively combined sections from multiple outputs to create a finalized document. Therefore, we consider each of the reconstructed briefs and articles in our output.

The average grades for the reconstructed articles and briefs were similar, with averages of 2.16 for technology specialists, 2.15 for marketing specialists, and 2.27 for web analysts. This result is somewhat intuitive given that the inputs to the tool were more formulaic than general use GenAI tools such as ChatGPT. It is, therefore, unsurprising that the output is similar between groups, since techniques such as prompt engineering skills were not as important for the bespoke tools.

Analysis with Employee-Defined Expertise

Per our pre-registration, we defined employee expertise based on the designations supplied by the AI team. To ensure robustness, we asked employees themselves during the pre-experiment survey the following question:

What best describes your current role?

Employees could select from the following options:

- SEO
- Market Analyst
- Marketing
- Developer
- Data Scientist
- Other [Click here to enter text]

We conducted a robustness check where we re-ran our analyses, defining web analysts as employees who answered "SEO", marketing specialists as employees who answered "Market Analyst" or "Marketing" and technology specialists as all other employees. While the expertise level supplied by the AI team usually aligned with the employee-defined expertise level, we found 10 employees who reported a different level of expertise.

Our results appeared to hold when using employee-defined expertise. The average brief grade without GenAI was 3.06, 3.08 and 3.63, and with GenAI was 4.04, 4.17 and 4.19, for technology specialists, marketing specialists and web analysts, respectively. Similarly, the average article grade without GenAI was 3.32, 2.86 and 3.38, and with GenAI was 3.45, 4.0 and 4.10, respectively.

GenAI Only

We requested, and received, 20 briefs and articles that were completed using the IG GenAI tools with minimal intervention from a member of the AI team. These articles and briefs can constitute a 'GenAI Only' benchmark.

Since the authors knew that these articles and briefs were created using GenAI, one of our MBA graders assessed these articles and briefs. The average score for briefs was 4.5, and the average score for articles was 3.89. Note that this is on par with — and for briefs, even better than — the articles and briefs completed by employees working with GenAI.

However, when we shared this result with the AI team, they confirmed that these tools were not meant to supplant, but merely complement, human workers. The articles for this experiment covered standard topics (trading gold futures and CFD commodities) that are largely robust to the news cycle; however, real articles released by IG should take into account evolving topics that best appeal to a contemporary reader base. For this reason — and others — IG emphasized that human employees are essential in this workflow. We conclude that, while impressive that the GenAI tool can achieve high performance with minimal intervention in this experiment, that performance may not be safely generalizable to practice.

APPENDIX B: ADDITIONAL MATERIAL

Deviation from Pre-Registration

In the pre-registration[9], we defined three groups — task experts, domain experts, and non-domain experts — as described in 3.2. We relabeled these groups as insiders, adjacent outsiders, and distant outsiders to more closely align with our theoretical framework.

Additionally, the pre-registration also references another form of expertise classification which, provided to us by IG, includes experts, moderate task experts, and novices and specifically refers to employee task expertise for each individual *task*. That is, the same employee might have different expertise classifications for different tasks brief and web article writing. We decided to focus our analysis on expertise at the profession level — web analysts, marketing specialists and other professions — for two main reasons. First, in discussions with IG they suggested that the profession-level classification was more important for internal organization; after all, employees are organized in departments, which center around specific roles. Second, our paper focuses on the potential blurring between professions, which is more appropriately analyzed with a profession-level classification of expertise.

We did not pre-register productivity as a dependent variable in 4.1. This was simply because we did not conceive of productivity as a dependent variable at the time of pre-registration —- it was suggested to us later by a colleague. However, productivity is a natural combination of two pre-registered dependent variables: output quality and completion speed.

Additionally, in our pre-registration, we defined three different task evaluators: ALPHA employees, research assistants, and LLM grades. For our analysis, we ended up using grades from four different evaluators: two authors and two MBA graders. We implemented several grading criteria post-hoc to ensure that all graders were providing sufficient quality grades.

First, all graders used the rubric developed by an ALPHA employee with extensive experience with these tasks. We were not able to receive evaluations from ALPHA on all responses due to resource constraints. However, the creator of the rubric provided grades for 10

[9] https://aspredicted.org/hx4b-mtsw.pdf

articles and 10 briefs. Any evaluator who provided grades for the subset of articles and briefs that on average deviated more than 2 points from the ALPHA employee grades was excluded. No evaluator failed this criteria. Second, following a short orientation session, all MBA evaluators were required to complete a short attention check to ensure they were familiar with the rubric and grading task. Both MBA graders answered all questions correctly. The attention check questions were as follows:

1. *What will you be grading?*
   - Blog posts and/or advertisements
   - Briefs and/or articles
   - Homework and/or essays
2. *What is the best possible score you can give?*
   - 5
   - 10
   - 100
3. *Which of these is NOT a criteria on the rubrics you will be using?*
   - Keywords
   - Meta Description
   - Number of quotes included from experts

We decided to exclude LLM grades from the final analysis. The research team used the OpenAI API to generate grades for all briefs and articles using the rubric developed by ALPHA. However, the resulting grades tended to be much higher on average with less variability than those generated by humans, with the majority of scores being 4s and 5s. This is consistent with other research, where LLMs tend to award higher grades than humans (Dell'Acqua et al., 2023). Given the inflated scores, we decided to exclude the LLM evaluations.

Finally, in our pre-registration, we described using soft skills to test heterogeneous treatment effects. The soft skills were to be collected and measured through a gamified test by a third-party. Unfortunately, due to challenges working with the third-party, we were unable to include these data in our analysis. Specifically, we were unable to verify the scientific rigor underlying these tests.

Evaluation Rubric

Brief scoring matrix

Factors to consider:

Keywords

How relevant are they – does content like the topic of the brief currently rank for that keyword on Google?

EG: 'how to trade gold futures' is highly relevant, 'gold futures' is somewhat relevant, and 'gold' is not relevant at all.

URL

It should include keywords relevant to the topic, and the URL should be in the /trading-strategies section

Meta description

Does it explain the topic succinctly and include a CTA within 160 characters?

Meta title and H1

Does it include main keywords and optimise for click-through?

EG: 'How to Trade Gold Futures: How to Get Started' ticks both boxes', 'How to Trade Gold Futures' is strong, and 'Gold Futures' is poor.

H2 headings:

Are there at least four H2 headings, including 2 covering 'What is', and 'How to'?

Survey Questions

Survey questions available upon request.

Interviews

TABLE B1
Interviews

ID	Pre-Experiment Interview	Post-Experiment Interview
1. AI Developers		
DEV_01	Y	Y
DEV_02	Y	Y
DEV_03	Y	Y
	3	3
2. Insiders		
TE_01	Y	N
TE_02	Y	N
TE_03	Y	Y
TE_04	N	Y
TE_05	N	Y
TE_06	N	Y
TE_07	N	Y
	3	5
3. Adjacent Outsiders		
DE_01	Y	N
DE_02	Y	N
DE_03	N	Y
DE_04	N	Y
DE_05	N	Y
DE_06	N	Y
DE_07	N	Y
	3	5
4. Distant Outsiders		
NO_01	Y	Y
NO_02	N	Y
NO_03	N	Y
NO_04	N	Y
NO_05	N	Y
	1	5
Total	9 Interviews	18 Interviews

Score	Summary
5	Highly relevant keywords, highly relevant URL, meta description includes succinct summary of topic and CTA in 160 characters, meta title and H1 including relevant keywords and optimised for click-through, at least four H2 headings including both 'What is' and 'How to'
4	Mostly relevant keywords, largely relevant URL, meta description includes either succinct summary of topic or CTA in 160 characters, meta title and H1 including either relevant keywords or click-through optimisation, at least four H2 headings including one of 'What is' or 'How to'
3	Some relevant keywords, somewhat relevant URL, meta description broadly describes the topic, meta title and H1 include somewhat relevant keywords, at least two H2 headings which are broadly relevant
2	Mostly irrelevant keywords, somewhat irrelevant URL, meta description doesn't describe the topic well, meta title and H1 do not do a good job of describing the topic and optimising for click-through, H2 headings are not highly relevant
1	Irrelevant keywords, irrelevant URL, meta description doesn't describe the topic, meta title and H1 do not include relevant keywords and do not optimise for click-through, H2 headings are not relevant and/or there are less than three

Article scoring matrix

Factors to consider:

Keywords

Includes the keywords that were briefed

Meta description

Does it explain the topic succinctly and include a CTA within 160 characters?

Meta title and H1

Includes one of the keywords that was briefed

H2 headings:

Are there at least four H2 headings, including 2 covering 'What is', and 'How to'?

Word count

Word count should be at least 1200 words

Internal links

Have they pointed to relevant internal links on ALPHA.com?

Readability and flow

Is the content readable, does it flow well sequentially, does it make use of bullet points and other methods to break up text, are paragraphs generally no longer than 50 words?

Grammar

Is the grammar and spelling accurate throughout?

User intent

Does the content provide a satisfying answer to the title of the piece?

Score	Summary
5	Includes all keywords that were briefed, meta description includes succinct summary of topic and CTA in 160 characters, meta title and H1 including relevant briefed keywords, at least four H2 headings including both 'What is' and 'How to', word count at least 1200 words, at least 3 relevant internal links, highly readable content, accurate spelling and grammar, aligns to user intent
4	Includes most keywords that were briefed, meta description includes either succinct summary of topic or CTA in 160 characters, meta title and H1 including relevant briefed keywords, at least four H2 headings including one of 'What is' or 'How to', word count at least 1000 words, at least 1 relevant internal link, readable content, accurate spelling and grammar, largely aligns to user intent
3	Includes some keywords that were briefed, meta description broadly describes the topic, meta title and H1 include some of the briefed keywords, at least two H2 headings which are broadly relevant, word count at least 800 words, somewhat readable content, somewhat accurate spelling and grammar
2	Doesn't include most of the keywords that were briefed, meta description doesn't describe the topic well, H2 headings are not highly relevant, content is difficult to read, spelling and grammar need improvement
1	Doesn't include most of the keywords that were briefed, meta description doesn't describe the topic, meta title and H1 don't include briefed keywords, H2 headings are not relevant and/or there are less than three, content is not readable, spelling and grammar is not accurate