© 2025 IJRTI | Volume 10, Issue 5 May 2025 | ISSN: 2456-3315

Use of Generative Artificial Intelligence

Leveraging Generative AI for Enhanced Computational Solutions

[1]Omkar Chavan, [2]Ritesh Uplaikar, [3]S.S. Balwante

[1]Student, [2]Student, [3]Professor
[1]School of Engineering,
[1]Ajeenkya DY Patil University, Pune, India
[1]omkarchavan2063@gmail.com, [2]riteshuplikar@gail.com, [3]shrikrishna.balwante@seamedu.com

***Abstract*—We study the staggered introduction of a generative AI-based conversational assistant using data from 5,172 customer support agents. Access to AI assistance increases worker productivity, as measured by issues resolved per hour, by 15% on average, with substantial heterogeneity across workers. Less experienced and lower-skilled workers improve both the speed and quality of their output while the most experienced and highest-skilled workers see small gains in speed and small declines in quality. We also find evidence that AI assistance facilitates worker learning and improves English fluency, particularly among international agents. While AI systems improve with more training data, we find that the gains from AI adoption are largest for relatively rare problems, where human agents have less baseline training and experience. Finally, we provide evidence that AI assistance improves the experience of work along two key dimensions: customers are more.**

Index Terms— Large Language Models, Worker ,Productivity, Generative AI, Technology Adoption.

I. Introduction (Heading 1)

Worker Learning, Experience of Work, Organizational Design. The emergence of generative artificial intelligence (AI) has attracted significant attention, but few studies have examined its economic impact. Although various generative AI tools have performed well in laboratory settings, excitement about their potential has been tempered by concerns that these tools may be less effective in real-world settings, where they may encounter unfamiliar problems, face organizational resistance, or provide misleading information in a consequential environment (Peng et al., 2023a; Roose, 2023).

In this paper, we study the adoption of a generative AI tool that provides conversational guidance to customer support agents. This is, to our knowledge, the first study of the impact of generative AI deployed at scale in the workplace. We find that access to AI assistance increases the productivity of agents by 15%, as measured by the number of customer issues they are able to resolve per hour. We find that these gains accrue disproportionately to less-experienced and lower-skill customer support workers. This finding suggests that generative AI systems may be capable of capturing and disseminating the behaviors of the most productive agents.

Computers and software have transformed the economy with their ability to perform certain tasks with far more precision,

polite and less likely to ask to speak to a manager.

speed, and consistency than humans. To be effective, these systems typically require explicit and detailed instructions for how to transform inputs into outputs: a software engineer must literally program the computer. Yet, despite significant advancements in traditional computing, many workplace activities—such as writing emails, analyzing data, or creating presentations—are difficult to "codify"—and have therefore defied automation.

Machine learning (ML) algorithms work differently from traditional computer programs: instead of requiring explicit instructions to function, these systems infer instructions from examples. Given a training set of images, for instance, ML systems can learn to recognize specific individuals even though one cannot fully explain what physical features characterize a given person's identity. This ability highlights a key distinguishing aspect of ML systems: they can learn to perform tasks even when no instructions exist—including tasks requiring tacit knowledge that could previously only be gained through lived experience (Polanyi, 1966; Autor, 2014; Brynjolfsson and Mitchell, 2017).

In addition, ML systems are often trained on data from human workers, who naturally vary in their abilities.
By seeing many examples of tasks—making sales pitches, driving a truck, or diagnosing a patient, to name a few— performed well and poorly, these models can implicitly learn what specific behaviors and characteristics set highperforming workers apart from the characteristics set high-performing workers apart from their less effective counterparts. That is, not only are generative AI models capable of performing complex tasks, they might also be capable of capturing the skills that distinguish top workers. The use of ML tools may therefore differentially expose lower-skill workers to new skills and techniques, leading to disparate changes in productivity even among workers performing the same task.

We study the impact of generative AI on productivity and worker experience in the customer service sector, an industry with one of the highest rates of AI adoption (Chui et al., 2021). We examine the staggered deployment of a chat assistant using data from 5,000 agents working for a Fortune 500 software firm that provides business process software. The tool we study is built on a recent version of the Generative Pre-trained Transformer (GPT) family of large language models developed by OpenAI (OpenAI, 2023). It monitors customer chats and provides agents with realtime suggestions for how to respond. It is designed to augment agents, who remain responsible for the conversation and are free to ignore or edit the AI's suggestions.
We have three sets of findings.

First, AI assistance increases worker productivity, resulting in a 15% increase in the number of chats that an agent successfully resolves per hour. This increase reflects shifts in three components of productivity: a decline in the time it takes an agent to handle an individual chat, an increase in the number of chats that an agent handles per hour (agents may handle multiple chats at once), and a small increase in the share of chats that are successfully resolved.

Within customer support workers, the productivity impacts of AI assistance are highly uneven. We find that less-skilled and less-experienced workers improve significantly across all productivity measures we consider, including approximate increase of 30% in the number of issues they are able to resolve per hour. Access to the AI tool helps newer agents move more quickly down the experience curve: treated agents with two months of tenure perform just as well as untreated agents with more than six months of tenure. In contrast, we find minimal impacts on the productivity of more-experienced or more-skilled workers. Indeed, we find evidence that AI assistance leads to a small *decrease* in the quality of conversations generated by the most skilled agents. These results contrast, in spirit, with studies that find evidence of skill-biased technical change for earlier waves

Our second set of results investigates the mechanism underlying our main findings. We show that AI recommendations appear useful to workers: agents who follow recommendations more closely see larger gains in productivity, and adherence rates increase over time, particularly those who were initially more skeptical. We also find that engagement with AI recommendations can generate durable learning. Using data on software outages—periods during which the AI system fails to provide suggestions—we find that workers maintain productivity gains relative to their pre-AI baseline even in the absence of AI recommendations. These improvements are particularly pronounced among agents with greater prior exposure to AI assistance, especially those who had been relying more on AIgenerated suggestions. Analyzing instances where the AI system was temporarily unavailable, we find that workers continue to perform above their pre-AI productivity levels, even without AI-generated suggestions. These productivity gains are more pronounced among agents who had previously engaged more with AI assistance, suggesting that exposure to AI contributes to lasting improvements in efficiency.

Examining AI's role across different types of customer inquiries, we observe that while AI systems benefit from larger datasets, their most significant value lies in assisting with uncommon issues—areas where human agents have limited prior experience. In contrast, agents already demonstrate proficiency in handling frequent problems, making AI's contribution in these cases less pronounced. Additionally, linguistic analysis of chat transcripts suggests that AI support enhances language fluency, particularly for international agents, and fosters a convergence in communication styles, with lessexperienced agents adopting patterns similar to their more skilled counterparts.

Beyond productivity and skill development, we assess how AI influences the overall work environment. Customer support roles can be challenging, often involving stressful interactions, overnight shifts, and outsourced work settings. While there is a concern that AI might make agent responses feel less personal, our findings suggest otherwise. AI assistance improves customer sentiment, leading to more respectful interactions and fewer escalations to supervisors. These shifts contribute to lower attrition rates, particularly among newer employees, indicating that AI can help create a more sustainable work environment. Overall, our findings highlight the potential of generative AI to enhance worker productivity and job satisfaction. However, these effects are observed within a single firm and over a limited period. The broader implications for labor markets remain uncertain. In the long term, firms may either increase hiring of entry-level agents due to productivity improvements or shift toward more advanced AI systems that reduce reliance on human labor. This raises critical questions about how AI will shape employment trends across different sectors in the future. Prior research suggests that technological advancements often provide greater benefits to higherskilled workers (Autor et al., 2003; Acemoglu & Restrepo, 2018; Acemoglu, 2024). However, our study does not capture changes in wages, overall labor demand, or shifts in the skill composition of newly hired employees, leaving these broader labor market effects outside our scope.

One of the key long-term challenges raised by AI adoption is its effect on worker incentives. In our findings, top-performing agents increasingly follow AIgenerated recommendations, despite a slight decline in the quality of their responses. This reliance on AI may reduce their original contributions, potentially limiting the ability of future AI models to adapt to new challenges. These findings raise important questions about whether and how workers should be compensated for the data they generate, which is crucial for improving AI systems.

Our research builds on a well-established body of work examining the impact of technological advancements on labor productivity and workplace organization (e.g., Rosen, 1981; Autor et al., 1998; Athey & Stern, 2002; Bresnahan et al., 2002; Bartel et al., 2007; Acemoglu et al., 2007; Hoffman et al., 2017; Bloom et al., 2014; Michaels et al., 2014; Garicano & Rossi-Hansberg, 2015; Acemoglu & Restrepo, 2020; Felten et al., 2023). Many of these studies, particularly those focusing on information technologies, suggest that IT adoption tends to benefit higher-skilled or more-educated workers (Akerman et al., 2015; Taniguchi & Yamada, 2022). For example, Bartel et al. (2007) found that IT adoption in manufacturing led to greater demand for skilled workers and increased skill requirements for machine operators. Similarly, Acemoglu & Restrepo (2020) explored the impact of robotics on employment and observed that job losses were most pronounced in blue-collar occupations and among workers without a college degree.

Despite extensive research on the effects of IT and automation, studies specifically examining AI-driven technologies remain limited. Some recent works, including Acemoglu et al. (2022), Zolas et al. (2020), and Calvino & Fontanelli (2023), analyze large-scale economic data from the US and OECD countries. Their findings indicate that AI adoption is more prevalent among large, high-productivity firms, particularly younger companies with the capacity to integrate new technologies. However, the evidence on AI's productivity impact remains mixed. Acemoglu et al. (2022) found no clear link between AI investments and productivity improvements, while Babina et al. (2022) reported a positive correlation between AI adoption and firm growth. These studies collectively highlight that the broader economic consequences of AI adoption remain uncertain, warranting further investigation.

effects of AI technologies may be challenging to identify at the macro-level because AI-adopting firms differ substantially from non-adopters.

Even at the decision level, the effects of AI on decision quality are mixed, often revealing unexpected challenges in human-AI collaboration. Some studies on AI decision support tools report positive impacts. For instance, Kanazawa et al. (2022) examine a non-generative AI tool for taxi drivers that suggests routes with the highest likelihood of finding customers. This tool reduces driver search time by 5%, with low-skill drivers seeing the largest reductions in search. On the other hand, several studies find that the combination of AI and human decisions performs worse than either AI or humans alone (Hoffman et al., 2017; Angelova et al., 2023; Agarwal et al., 2023; Poursabzi-Sangdeh et al., 2021). In fact, a metaanalysis of more than 100 experimental studies concludes that, on average, human-AI collaborations underperform both the AI alone and the best human decision-makers (Vaccaro et al., 2024). These results underscore the particular challenges introduced when using AI-based tools designed to augment human decision making.

1. Generative AI and Large Language Models The rapid advancement of artificial intelligence, along with the public release of tools like ChatGPT, GitHub Copilot, and DALL-E, has sparked widespread discussion—ranging from excitement to concern (The White House, 2022). These systems belong to a broader category of machine learning known as generative AI, which can create original content such as text, images, music, and videos by recognizing and replicating patterns found in existing data. This section provides an overview of generative AI technology and explores its potential economic implications.

1.1 Technical Overview

A key subset of generative AI is large language models (LLMs), which are deep learning systems designed to process and generate sequential data (Bubeck et al., 2023). These models are trained on vast text corpora— such as Wikipedia, digitized books, and portions of the internet—by predicting the next word in a sequence based on prior context. This process enables them to produce fluent, meaningful text. While LLMs are primarily designed for natural language processing, the same techniques can be applied to other structured sequences, including protein chains, audio, programming code, and even chess moves (Eloundou et al., 2023). The rapid improvements in generative AI capabilities have been driven by several key factors: scaling computational power, advancements in model architecture, pre-training on extensive datasets, and refinements in training methodologies.

Computational Scaling

The effectiveness of LLMs is strongly linked to the scale of computation, including the volume of training data, the number of model parameters, and the computational power used (Kaplan et al., 2020). As AI companies continue to expand these resources, the size and complexity of models have grown dramatically. For example, GPT-3 was trained with 175 billion parameters on 300 billion tokens, incurring computing costs estimated at $5 million. Its successor, GPT-4, is believed to have approximately 1.8 trillion parameters and was trained on 13 trillion tokens, with computational expenses reaching an estimated $65 million (Li, 2020; Brown et al., 2020; Patel & Wong, 2023).

Advances in Model Architecture

Modern LLMs build on foundational innovations such as positional encoding and self-attention mechanisms. Positional encoding ensures that the model retains information about word order within a given text input. Meanwhile, self-attention assigns varying importance to words based on the broader context of the text, allowing models to capture long-range dependencies and semantic relationships, even when processing text in parallel (Vaswani et al., 2017; Bahdanau et al., 2015). These innovations have significantly improved the ability of LLMs to generate coherent, contextually aware responses Next, LLMs can be pre-trained on large amounts of unlabeled data from sources such as Reddit or Wikipedia. Because unlabeled data are far more prevalent than labeled data, LLMs can learn about natural language on a much larger training corpus (Brown et al., 2020). By seeing, for example, that the word “yellow” is more likely to be observed with “banana” or “sun” or “rubber duckie,” the model can learn about semantic and grammatical relationships even without explicit guidance (Radford and Narasimhan, 2018). The resulting model can be used in multiple applications
because its training is not specific to a particular set of tasks.
General-purpose LLMs can be further refined through fine-tuning, allowing them to adapt to specific applications and priorities (Ouyang et al., 2022; Liu et al., 2023). For instance, a model designed for social media content generation could be enhanced by incorporating labeled data that not only include posts or tweets but also indicate engagement metrics such as likes and shares. Similarly, while an LLM may generate multiple possible responses to a query, some may contain inaccuracies or inappropriate language. To improve reliability, human reviewers can rank these responses, helping to train a reward function that prioritizes accurate and appropriate outputs. These targeted refinements enable general-purpose models to perform more effectively in specialized contexts (Ouyang et al., 2022).
Collectively, these advancements have led to significant improvements in AI model performance. The Generative Pre-trained Transformer (GPT) family, in particular, has gained widespread attention for its rapidly evolving capabilities and expanding applications.

1.2 The Economic Impacts of Generative AI Traditionally, computers have been highly effective at executing predefined instructions, making them wellsuited for tasks that follow clear, rule-based processes (Autor, 2014). As a result, automation has significantly reduced the need for workers in roles centered around routine tasks, such as bookkeeping, data entry, and assembly line operations, leading to wage declines in these fields (Acemoglu & Autor, 2011).
At the same time, technological advancements have increased demand for workers with complementary skills, including programming, data analysis, and research. These shifts, alongside changes in the supply of skilled labor, have contributed to growing wage inequality in the United States and have influenced broader organizational transformations (Katz & Murphy, 1992; Autor et al., 2003; Michaels et al., 2014; Bresnahan et al., 2002; Baker & Hubbard, 2003; OECD, 2023).
Unlike traditional software that relies on predefined instructions, generative AI can perform tasks without needing explicit programming for each scenario. For instance, if prompted to compose an email declining a raise request, an AI model is likely to generate a response that is professional and empathetic. This occurs because the model has been trained on a vast dataset containing workplace communications, where similar requests have been denied using such language. Notably, the model does this without a programmer explicitly defining what a "professional" or "conciliatory" tone should be. In fact, the ability to navigate social and professional norms appropriately is often intuitive and difficult to codify— even for those who possess it. This kind of implicit understanding, often referred to as tacit knowledge, underlies many of the tasks humans perform in both professional and everyday settings (Polanyi, 1966; Autor, 2014).

The ability of generative AI to absorb and apply tacit knowledge suggests that these systems can perform tasks traditionally requiring human judgment and experience. This expands the range of activities that machines can assist with, including non-routine tasks that were previously considered beyond automation's reach. For example, GitHub Copilot, an AI tool for software development, not only provides intelligent code suggestions but can also explain the logic behind its outputs in natural language (Nguyen & Nadi, 2022; Zhao, 2023). These advancements are extending AI's role into fields requiring significant expertise, such as mathematics, scientific analysis, financial modeling, and software engineering. Since these areas have traditionally been dominated by highly skilled professionals—who have often benefited from technological progress rather than being displaced by it—the rise of generative AI may significantly reshape the relationship between automation, labor demand, and economic inequality (The White House, 2022).

Beyond expanding the scope of automation, generative AI can also provide insights into the behaviors that differentiate top-performing workers from others. Since these models are trained on vast amounts of humangenerated data, they learn from both effective and ineffective examples of task performance. By recognizing patterns that contribute to success, AI systems can implicitly identify best practices and replicate them in new contexts. This capability presents organizations with multiple strategic options: they may use AI to replace lower-skilled workers, deploy AI-driven training to help employees improve, or enable less experienced workers to adapt more quickly by learning from high-performing counterparts. Regardless of the approach, the impact of generative AI is unlikely to be uniform—its effects will likely vary based on worker skill levels, even among those performing the same tasks.

Despite their potential, generative AI tools face significant challenges in real-world applications. At a technical level, popular LLM-based tools, such as ChatGPT, have been shown to produce false or misleading information unpredictably, raising concerns about their reliability in high-stakes situations. Second, while LLM models often perform well on specific tasks in the lab (OpenAI, 2023; Peng et al., 2023b; Noy and Zhang, 2023), the types of problem that workers encounter in real-world settings are likely to be broader and less predictable. Furthermore, generative AI tools often require prompts from human operators, yet finding ways to effectively combine human and AI expertise may be challenging: for instance, earlier research indicates that decision-support systems integrating AI with human judgment often perform worse than those that rely on the human or the AI alone (Vaccaro et al., 2024). These challenges raise concerns about the ability of AI systems to provide accurate assistance in every circumstance and—perhaps more importantly—workers' capacity to distinguish cases where AI suggestions are effective from those where they are not. Finally, the efficacy of new technologies is likely to depend on how they interact with existing workplace structures. Promising technologies may have more limited effects in practice due to the need for complementary organizational investments, skill development, or business process redesign. Because generative AI technologies are only beginning to be used in the workplace, little is currently known about their impacts.

2. Setting: LLMs in Customer Support

2.1 Generative AI in Customer Support

The customer service industry has become one of the leading adopters of generative AI, leveraging it to enhance efficiency and improve interactions with customers. Effective customer support is essential for maintaining strong client relationships and upholding a company's reputation. However, as in many fields, there is significant variation in worker productivity (Berg et al., 2018; Syverson, 2011).

New employees often require extensive training and take time to reach peak productivity. At the same time, employee turnover in this sector is exceptionally high— industry estimates indicate that around 60% of contact center agents leave their positions annually. This attrition results in significant costs for companies, with expenses ranging from $10,000 to $20,000 per departing employee (Buesing et al., 2020; Gretz & Jacobson, 2018). To mitigate these challenges, supervisors typically dedicate at least 20 hours per week to coaching lowerperforming agents (Berg et al., 2018). Given these persistent issues—fluctuating productivity, high turnover, and substantial training investments— companies are increasingly integrating AI-driven solutions to support and streamline customer service operations (Chui et al., 2021).

At a technical level, customer support is well-suited for current generative AI tools. From an AI's perspective, customer-agent conversations can be thought of as a series of pattern-matching problems in which one is looking for an optimal sequence of actions. When confronted with an issue such as "I can't login," an AI/agent must identify which types of underlying problems are most likely to lead a customer to be unable to log in and think about which solutions typically resolve these problems ("Can you check that caps lock is not on?"). At the same time, they must be attuned to a customer's emotional response, making sure to use language that increases the likelihood that a customer will respond positively ("that wasn't stupid of you at all! I always forget to check that too!"). Because customer service conversations are widely recorded and digitized, pre-trained LLMs can be fine-tuned for customer service using many examples of both successfully and unsuccessfully resolved conversations.

Customer service is an industry characterized by significant variation in the skills and effectiveness of individual agents. High-performing customer support representatives tend to excel at identifying the root cause of a customer's issue based on their description. These top agents typically ask more clarifying questions before diagnosing the problem, which may initially take more time but ultimately prevents wasted effort on incorrect solutions. Such behavioral patterns can often be identified through the extensive training data available to AI models designed for customer service. As a result, generative AI has the potential to capture and replicate the "best practices" demonstrated by the most skilled agents.

The following sections will provide an overview of the company analyzed in this study and the AI tool it implemented.

2.2 Data Firm Background

For this study, we partnered with a company that develops AI-driven customer service solutions (referred to as the "AI firm") to examine the impact of their tool within one of their client organizations (referred to as the "data firm").

The data firm is a Fortune 500 company specializing in business process software designed for small and medium-sized enterprises in the United States. It employs a combination of in-house customer support agents and outsourced personnel from third-party service providers. Most of the agents in our dataset are based in the Philippines, with a smaller segment working in the United

States and other countries. Despite their geographical distribution, all agents perform a similar role: assisting U.S.-based small business owners with technical support inquiries.

Within this group of customer service agents, chat assignments were based solely on availability, without any pre-screening. Support sessions were relatively lengthy, averaging about 40 minutes, as a significant portion of the conversation was dedicated to diagnosing the core technical issue.

The role demands a combination of in-depth product knowledge, problem-solving abilities, and the capacity to handle interactions with frustrated customers. The company evaluates agent productivity using three key industry-standard metrics:

Average Handle Time (AHT): The average duration required for an agent to complete a customer chat. Resolution Rate: The percentage of conversations where the agent successfully resolves the customer's issue. Net Promoter Score (NPS): A customer satisfaction metric derived from post-chat surveys, representing the percentage of customers who would recommend the agent minus those who would not.

A highly productive agent is expected to efficiently manage customer interactions while maintaining a high resolution rate and strong customer satisfaction scores. Regardless of their location, agents are structured into teams led by managers who provide ongoing feedback and training. Weekly one-on-one sessions between managers and agents serve as an opportunity to discuss solutions to emerging software issues, explain regulatory updates, or offer strategies for handling customer frustration. Agents work independently, and their performance does not directly influence that of their colleagues.

Compensation typically includes a base hourly wage, with additional performance-based incentives tied to metrics like chat resolution rate and the number of chats handled per hour. While specific salary data is unavailable, interviews with managers suggest that performance bonuses make up approximately 20% to 40% of an agent's total earnings.

2.3 AI System Design

The AI system under study is designed to recognize conversational patterns that contribute to efficient issue resolution. It is built on a recent version of GPT and finetuned using a large dataset of past customer-agent interactions, labeled with key outcomes such as successful problem resolution and chat duration. For training purposes, the AI system also incorporates data on whether an agent was classified as a "top" performer by the company. Many elements of effective agent performance, such as asking strategic clarifying questions, demonstrating attentiveness, de-escalating tense situations, and adapting responses to customer concerns, are difficult to quantify. However, the AI system is trained to identify and replicate these successful conversational behaviors to improve customer interactions.

The AI model is explicitly trained to recognize and replicate effective communication styles, simplify complex concepts, and maintain a professional tone. By learning from top-performing agents, the system is designed to incorporate subtle yet impactful behaviors, such as demonstrating empathy, adapting responses to customer sentiment, and using clear, concise explanations.

Additionally, the AI firm refines its model using a process similar to that of Ouyang et al. (2022), ensuring that generated responses prioritize key elements like empathy, appropriate technical references, and professional language. This targeted training helps mitigate common concerns related to using large language models (LLMs) for customer interactions. Once integrated into the workflow, the AI system delivers two primary types of assistance:

Real-time response suggestions – The AI provides agents with suggested replies based on the context of the conversation.

Technical documentation links – The system suggests relevant internal resources to help agents resolve customer inquiries efficiently.

Both types of recommendations are generated in realtime based on the ongoing exchange between the customer and the agent.

An example of AI assistance is illustrated in Appendix Figure A.1. In the chat window (Panel A), a customer named Alex describes their issue. The AI system then suggests two potential responses (Panel B), incorporating customer-friendly phrasing such as "I can definitely assist you with this!" and "Happy to help you get this fixed asap," which have been correlated with positive outcomes. Additionally, when a technical issue is detected, the AI can recommend a relevant internal resource (Panel C), directing the agent to the appropriate documentation.

Crucially, this AI system is designed to support rather than replace human agents. All AI-generated suggestions are visible only to the agent, who retains full control over whether to use, modify, or disregard them. This humanin-the-loop approach reduces the risk of incorrect or irrelevant outputs being relayed to customers. Moreover, the AI system does not generate suggestions when it encounters unfamiliar situations with insufficient training data, ensuring that agents rely on their expertise when necessary.

3. Empirical Strategy, Data, and Deployment. This section outlines the implementation of the AI system, the data collected for analysis, and the methodology used to evaluate its impact.

3.1 AI Rollout

The deployment of AI assistance began with a small-scale randomized pilot involving a limited number of agents. The full rollout took place between Fall 2020 and Winter 2021, with timing influenced by two key constraints: onboarding bottlenecks due to limited training resources and budgetary restrictions on AI adoption. Agents could access the AI tool only after completing a structured three-hour online training session, conducted by the AI firm. To ensure quality and consistency, these sessions were kept small and led by two experienced employees from the AI firm, both of whom had prior contact center experience and extensive knowledge of the AI system. Given their other responsibilities, these trainers could only offer a limited number of sessions each week, and session schedules were adjusted to accommodate the global workforce across different time zones.

Due to the relatively high cost and untested nature of generative AI at that time, the data firm allocated a restricted budget for its deployment. Once the number of onboarded agents reached the contractual limit, further enrollments were halted. However, if an AIenabled agent left, new agents were added under the same licensing agreement. These constraints in licensing and training capacity resulted in a staggered adoption process, which serves as the foundation for our analysis. Managers at each office were responsible for assigning agents to training sessions. Interviews with AI firm employees suggest that managers made these decisions strategically to minimize disruptions to customer service. Since enrolling an entire team in a single training session would have significantly increased customer wait times, managers distributed team members across different sessions to maintain service continuity. Following the initial onboarding session, no additional training on AI usage was provided, as the AI firm had a small product management team and lacked the capacity to offer ongoing support to the thousands of agents using the tool.

As a result, AI adoption varied at the individual level within the same team and office. In October 2020, only around 5% of active team members had access to AI assistance, but this figure increased to approximately 70% by January 2021. While our primary analysis focuses on individual adoption timelines, we also present supplementary results in Appendix Table A.2, where team-level adoption patterns are used as an instrument for individual adoption dates.

3.2 Summary Statistics

Table 1 presents an overview of the sample characteristics, categorized into four groups: all agents ("all"), those who never received access to the AI tool during the study period ("never treated"), observations before AI access for agents who eventually adopted it ("treated, pre"), and observations after AI access ("treated, post").

Our dataset includes approximately 3 million customer interactions from 5,172 agents. Of these, 1.2 million chats were handled by 1,636 agents during the post-AI adoption period. A majority of the agents in our dataset—89%—are based outside the United States, primarily in the Philippines. For each agent, we track their assigned manager, tenure, geographical location, and whether they are employed directly by the data firm or through a subcontractor.

To evaluate the impact of AI deployment, we use key performance metrics, aggregated at the agent-month level—the most detailed level with complete data. Our primary measure of productivity is resolutions per hour (RPH), which represents the number of chats an agent successfully resolves per hour. Since RPH is influenced by multiple factors, we also analyze these separately:

Average Handle Time (AHT): The average duration of a chat.

Chats Per Hour (CPH): The number of chats an agent manages per hour, accounting for multitasking.

Resolution Rate (RR): The proportion of chats successfully resolved.

Net Promoter Score (NPS): A customer satisfaction metric derived from post-chat surveys.

While our primary outcome metrics are calculated at the agent-month level, some data, such as chat duration, is available at a more granular level for individual chats. We also derive additional insights from chat transcripts, including sentiment analysis, topic categorization, and language fluency.

AHT and CPH are available for all agents in our dataset, but subcontracted agents do not always have consistent call quality records. Consequently, our core productivity metric, resolutions per hour (RPH), is only available for the subset of agents with complete call quality data. Since some agents work only part-time or for a portion of the year, we calculate year-month observations based only on periods when an agent actively handled chats. A more detailed discussion of data construction and key variables can be found in Appendix Section A.2. Figure 1 illustrates the raw distributions of key performance outcomes across the different groups. Panels A through D show that agents with AI assistance demonstrate improved performance across multiple metrics compared to both never-treated agents and their own pre-AI performance. However, Panel E indicates no significant differences in customer satisfaction (NPS) between AI-enabled and non-AI agents.

Focusing on our main productivity measure, Panel A of Figure 1 and Table 1 show that nevertreated agents resolve an average of 1.7 chats per hour, whereas posttreatment agents resolve 2.5 chats per hour. Some of this difference may be due to initial selection: treated agents have higher resolutions per hour prior to AI model deployment (2.0 chats) relative to never treated agents (1.7). This same pattern appears for chats per hour (Panel C) and resolution rates (Panel D): while evertreated agents appear to be stronger performers at the outset than agents who are never treated, post-treatment agents perform substantially better. Looking instead at average handle times (Panel B), we see a starker pattern: pre-treatment and never-treated agents have similar distributions of average handle times, centered at 40 minutes, but post-treatment agents have a lower average handle time of 35 minutes. These figures, of course, reflect raw differences that do not account for potential confounding factors such as differences in agent experience or differences in selection into treatment. In the next section, we will more precisely attribute these raw differences to the impact of AI model deployment.

3.3 Empirical Strategy

We isolate the causal impact of access to AI recommendations using a standard difference-in- differences regression:

$$y_{it} = \delta_t + \alpha_i + \beta AI_{it} + \gamma X_{it} + \epsilon_{it} \quad (1)$$

Our outcome variables, y_{it}, is performance measures for agent i in year-month t, with resolutions per hour as our main measure of productivity. We measure these outcomes in levels, and report percentage changes off the baseline pre-treatment means. Our main variable of interest is AI_{it}, an indicator that equals one if AI assistance is activated for agent i at time t. All regressions include year-month fixed effects, δ_t, to control for common, time-varying factors such as tax season or the end of the business quarter. In our preferred specifications, we also include agent fixed effects α_i to control for time-invariant differences in productivity across agents and time-varying tenure controls X_{it} (specifically, fixed effects for agent tenure in months). In our main specifications, we weight each agent-month equally and cluster standard errors at the agent level to reflect the fact that AI access is rolled out individually, but Appendix Tables A.3 and A.4 show that our results are robust to alternative weightings and clustering.

A rapidly growing literature has shown that two-way fixed effects regressions deliver consistent estimates only with strong assumptions about the homogeneity of treatment effects and may be biased when treatment effects vary over time or by adoption cohort (Cengiz et al., 2019; de Chaisemartin and D’Haultfœuille, 2020; Sun and Abraham, 2021; Goodman-Bacon, 2021; Callaway and Sant’Anna, 2021; Borusyak et al., 2022). For example, workers may take time to adjust to using the AI system, in which case its impact in the first month may be smaller. Alternatively, the onboarding of later cohorts of agents may be smoother, so that their treatment effects may be larger.

We study the dynamics of treatment effects using the interaction weighted (IW) estimator proposed in Sun and Abraham (2021). Sun and Abraham (2021) show that this estimator is consistent assuming parallel trends, no anticipatory behavior, and cohort-specific treatment effects that follow the same dynamic profile. Both our main differences-in-differences and event study estimates are similar using robust estimators introduced in de Chaisemartin and D'Haultfœuille (2020), Borusyak et al. (2022), Callaway and Sant'Anna (2021), and Sun and Abraham (2021), as well as using traditional two-way fixed effects OLS. In fact, Appendix Figure A.10 shows similar treatment effects across adoption cohorts (e.g. those that received AI access earlier or later, and were thus subject to potentially different versions of the model). We also show our results our similar clustering at different levels of granularity (Appendix Table A.3) and instrumenting agent adoption with the date the first worker within the agent's team received AI access (Appendix Table A.2).

4. Topic Handling, Learning, Adherence, and Conversational Change

In this section, we analyze different factors influencing the impact of AI assistance on agent performance. We begin by assessing how agents interact with AIgenerated suggestions. On average, agents incorporate AI recommendations 35% of the time, demonstrating a selective approach to AI guidance. Our findings show that those who consistently follow AI recommendations experience the greatest performance improvements. Over time, adherence rates increase across all workers, with more experienced agents showing the most significant rise. By the end of our study, adherence levels are fairly uniform across employees, regardless of their skill level or tenure, indicating a growing trust in AIgenerated suggestions.

Next, we evaluate whether AI exposure leads to lasting skill improvements. By examining software outages— periods when AI assistance was temporarily unavailable—we find that agents who had prior AI exposure continue to perform better even when AI support is removed. This effect is more pronounced for agents with longer exposure to AI assistance and those who frequently relied on AI suggestions while the system was operational.

Our findings suggest that AI significantly aids agents in handling complex or uncommon customer issues by providing relevant information in real time. Additionally, AI support helps enhance agents' English language proficiency, with the greatest improvements observed among those with lower initial fluency. This indicates that AI assistance not only improves technical accuracy but also refines overall communication effectiveness.

To further analyze the impact of AI, we examine how agents' conversation styles evolve post-AI adoption. We find that lower-skilled agents experience more significant shifts in language use, suggesting that AI support helps standardize communication practices. Additionally, conversations conducted by lower-skilled agents become more aligned with those of higher-skilled colleagues, implying that AI helps bridge skill gaps and promotes consistency in customer interactions.

4.1 Adherence to AI recommendations The AI tool we study makes suggestions, but agents are ultimately responsible for what they say to the customer. Thus far, our analysis evaluates the effect of AI assistance, irrespective of the frequency with which users adhere to its suggestions. Here, we examine how closely agents adhere to AI recommendations, and document the association between adherence and returns to adoption. We define "adherence" as the proportion of AI suggestions an agent typically adopts. The AI company considers an agent to have adhered when they either directly copy the AI's proposed text or manually enter highly similar content. To gauge initial adherence, we classify each treated agent
into a quintile based on their level of adherence during their first month using the AI tool.

The analysis of AI adherence among customer service agents reveals several important insights into how AIgenerated recommendations are used and how they impact productivity.

Understanding AI Adherence and Its Impact Agents do not accept AI-generated recommendations blindly; instead, they selectively integrate them into their workflow. On average, agents incorporate 38% of AI suggestions, with adherence rates ranging between 23% and 50%. This indicates that while AI assistance can be beneficial, agents frequently disregard recommendations—possibly because some are irrelevant or inaccurate. Similar adherence rates have been observed in other AI-assisted tools like GitHub Copilot, where software developers accept between 27% and 46% of AI-generated code suggestions (Zhao, 2023). This behavior aligns with the reality that AI models, despite their capabilities, do not always provide optimal solutions.

The variation in adherence is consistent across different locations and teams, suggesting that it is not influenced by organizational culture or managerial policies but rather by individual agent preferences and needs.

The Relationship Between Adherence and Productivity the extent to which an agent follows AI recommendations is strongly linked to their productivity gains. Agents were categorized into different groups based on their adherence levels in the first month of AI access.

The data reveals a clear pattern:

Even agents with the lowest adherence levels still experienced around a 10% increase in productivity after AI adoption. Agents who followed AI recommendations the most—falling in the highest adherence category— saw productivity gains of nearly 25%, more than twice the improvement observed in the lowest adherence group.

This suggests that following AI recommendations contributes to better performance, but the degree of benefit depends on how frequently agents utilize AI guidance. The strongest productivity improvements were observed in metrics such as average handle time (AHT) and chats per hour (CPH), while the effects on resolution rate (RR) and customer satisfaction (NPS) were more variable.

The correlation between adherence and productivity could be influenced by multiple factors:

Self-Selection Bias: Agents who are naturally more productive may be more inclined to use AI recommendations effectively.

Selection on Gains: Some agents may follow AI suggestions more frequently because they derive greater benefits from them.

To explore whether these effects are due to selection or actual learning, the study examines how adherence changes over time. If adherence were purely based on initial preference, agents with low adherence would continue to ignore AI recommendations. However, the data shows that even those who initially followed AI recommendations less frequently increased their adherence over time. This suggests that AI assistance provides real learning benefits, leading agents to gradually adopt more AI-generated suggestions as they recognize their usefulness.

Workers who initially exhibit low adherence to AI recommendations tend to increase their compliance over time. For example, those in the bottom adherence group initially follow AI recommendations less than 20% of the time but, within five months, their adherence increases by over 50%, reaching a compliance rate of just over half.
Adherence trends also vary based on worker tenure at the time of AI deployment. More experienced workers are initially less likely to follow AI recommendations, with adherence rates around 30% for those with over a year of experience compared to 37% for those with less than three months. However, over time, all workers tend to increase their adherence, with more tenured employees adopting AI recommendations at a faster rate. Within five months, adherence levels converge across experience levels.
When analyzed by skill level, adherence rates at the start are relatively similar across different proficiency groups. Regardless of their initial expertise, all skill levels show a steady increase in AI compliance over time. This pattern suggests that adherence is not solely driven by selection bias (where only certain types of workers adopt AI), but rather by a growing recognition of AI's usefulness over time.

POTENTIAL OVER-RELIANCE ON AI

While greater adherence generally correlates with productivity gains, there is an interesting trend among high-skill agents. Despite their relatively smaller productivity improvements—and, in some cases, even a decline in certain quality measures—these agents also increase their AI adherence at a rate similar to their lower-skill peers.
This raises the possibility that some workers may be overrelying on AI recommendations, even when doing so is not necessarily optimal. High-performing agents may default to AI-generated responses rather than taking the time to provide higher-quality, customized inputs. If AI recommendations become a substitute for their expertise rather than a complement, it could reduce the diversity and quality of human-generated solutions, which are essential for improving future AI training. However, no evidence suggests that AI model quality is declining over time. Workers who gained access to later, more updated versions of the AI system experienced similar productivity boosts in their first month as those who started with earlier versions. This suggests that, at least in the short term, AI-assisted workflows are not deteriorating in effectiveness despite increasing reliance on the model.

4.2 Worker Learning and AI Exposure

A key question remains: Are the observed productivity improvements and communication changes temporary adjustments to AI assistance, or do they reflect long-term enhancements in workers' skills?
Understanding whether these behavioral shifts lead to durable skill development is crucial. If AI merely fills knowledge gaps temporarily, then productivity improvements might vanish when AI assistance is removed. However, if workers internalize AI-driven best practices, their performance should remain high even when AI is unavailable.
By examining instances where AI assistance was temporarily disabled due to software outages, researchers found that workers who had prolonged exposure to AI continued to perform better even when the AI was not available. This suggests that interacting with AI recommendations contributes to long-term skill development, particularly in problem-solving, communication, and workflow efficiency. Overall, AI not only boosts immediate productivity but also acts as a learning tool, helping workers refine their expertise and close skill gaps over time. However, striking the right balance between leveraging AI guidance and maintaining independent critical thinking will be essential to ensuring long-term benefits.
A key question raised by our findings so far is whether these improvements in productivity and changes in communication patterns reflect durable changes in the human capital of workers or simply their growing reliance on AI assistance. In the latter case, the introduction of AI assistance could actually lead to an erosion in human capital, and we would expect treated workers to be less able to address customer questions if they are no longer able to access AI assistance.
To study this, we examine how workers perform during periods in which they are not able to access AIrecommendations due to technical issues at the AI firm. Outages occur occasionally in our data and can last anywhere from a few minutes to a few hours. During an outage, the system fails to provide recommendations to some, but not necessarily all, workers. For example, outages may affect agents who log into their computers after the system crashes, but not agents working at the same time who had signed in earlier. They may also affect workers using one physical server but not another. Our AI firm tracks the most significant outages in order to perform technical reviews of what went wrong. We compile these system reports to identify periods in which a significant fraction of chats are affected by outages.

4.3 Non-Routine Topics and Handling Routine with AI Assistance

So far, the analysis has focused on how the impact of AI assistance varies based on worker characteristics, such as skill level, tenure, and adherence. However, the effectiveness of AI support also depends on the type of customer problems it is designed to address. Customer service agents handle a wide variety of inquiries, ranging from routine and well-documented issues (such as resetting passwords or onboarding employees) to rare and complex cases (such as ensuring compliance with international tax regulations or managing wage garnishments).
Since AI models are trained on historical customer interactions, they tend to have more exposure to frequently encountered issues. This raises an important question: Can AI provide useful recommendations for less common problems, or does its effectiveness decline as inquiries become more unique?

AI's Impact on Different Types of Customer Issues To examine this, customer interactions can be categorized based on how common or rare the issues are. As expected, a small number of common topics dominate customer service interactions, while a long tail of rarer issues makes up the rest. For instance, the majority of conversations tend to revolve around payroll, taxes, and account management, while more specialized inquiries—such as complex financial transactions or compliance issues—occur less frequently. The effectiveness of AI support varies depending on how often a particular type of issue arises. Interestingly, the impact of AI on productivity is not uniform across different categories of customer problems. The largest productivity improvements are observed not in the most routine tasks, nor in the most rare ones, but in moderately uncommon issues.

For highly routine problems, AI reduces response times only slightly. This is because customer service agents— especially those with some experience—are already well-trained in handling these issues efficiently. For example, resetting a password or updating an account detail is a task that even novice agents can perform quickly, meaning AI assistance adds little marginal benefit.

For moderately rare problems, AI assistance has the greatest impact. These are issues that arise often enough for the AI model to have been trained on relevant examples, yet they may not be common enough for human agents to have extensive experience handling them. In such cases, AI can significantly reduce resolution time by providing relevant documentation, suggested responses, or troubleshooting guidance, reducing the need for agents to manually search for solutions. For very rare problems, AI assistance still provides benefits, but the impact is smaller compared to moderately uncommon cases. Since these issues occur infrequently, the AI system has had less exposure to similar interactions during training, making its recommendations less reliable. At the same time, agents themselves may also lack expertise in these areas, leading to limited productivity improvements even when AI assistance is available.

AI'S COMPLEMENTARY ROLE

These findings highlight an important distinction between the technical quality of an AI model and its realworld impact on productivity. AI systems generally perform best when trained on large, diverse datasets, allowing them to develop robust patterns and reduce errors. However, when AI is used to assist human workers, its effectiveness is not solely determined by its training quality but also by how well it complements human expertise.

In areas where agents are already highly skilled, AI assistance may not provide much additional value. Conversely, in situations where AI has too little training data, its recommendations may be unreliable. The greatest productivity benefits arise when AI is used to bridge gaps in human expertise—providing support in areas where agents may lack direct experience but where AI has seen enough relevant cases to offer reliable guidance.

For example, AI assistance can be particularly valuable in cases where agents would otherwise have to pause a conversation to search for information. By instantly suggesting relevant documentation or recommended actions, AI can reduce customer wait times and improve overall efficiency.

BALANCING AI ASSISTANCE AND HUMAN EXPERTISE

Ultimately, the effectiveness of AI in customer service depends on its ability to enhance human capabilities rather than replace them. AI is most beneficial when used to supplement areas where agents have limited experience, rather than in tasks where they are already proficient or where AI itself lacks sufficient knowledge. By strategically deploying AI assistance where it has the greatest impact, organizations can optimize customer service workflows, ensuring that AI supports human agents rather than diminishing their ability to solve problems effectively.

4.4 Conversational Style

Our previous analysis examined how the effectiveness of AI assistance varies based on the technical nature of the problems being addressed. However, top customer service agents not only resolve technical issues but also communicate in a way that ensures customer satisfaction. For our USbased customers, this involves clearly writing in fluent English and displaying subtle interpersonal skills. In this section, we first demonstrate that AI assistance improves overall language fluency. Then, to capture less quantifiable changes in communication style, we compare the evolution of conversational text produced by different workers. Our analysis provides suggestive evidence that AI adoption helps lower-skill workers communicate more like their higher-skill peers.

4.4.1 English Fluency For contact center workers serving US customers, the ability to communicate in clear, idiomatic

English is crucial to their job performance and customer satisfaction. We begin by assessing how AI assistance influences workers' ability to communicate clearly. In our data, 80% of the agents are based in the Philippines, where many residents are fluent English speakers. However, cultural differences and language nuances can occasionally lead to misunderstandings or a sense of disconnect, even when an agent's technical language skills are strong. We measure the proficiency of text in two ways: its comprehensiblity and its native fluency. The comprehensibility score assesses whether the agent produces text that is cogent and easy to understand, using a scale of 1 to 5, where 1 indicates "very difficult to comprehend" and 5 signifies "very fluent and easily understandable, with no significant errors." In contrast, our native fluency

focuses on whether the text was likely to have been produced by a native speaker of American English. Our criteria for native-like fluency are based on the Interagency Language Roundtable "functionally native" proficiency standard. This is also a 5 point scale where 1 indicates that a writer is "Definitely not a native American English speaker" and 5 indicates that they definitely are. For instance, "I could care less" is grammatically incorrect, but a common English-language expression. On the other hand, Filipino agents often use the greeting "to have a blessed day," which is grammatically correct, but not a common greeting in the United States. We use Gemini, an LLM, to score agents' text in each conversation. For more information on our specific approach, prompts, and validation tests

4.4.2 Textual Convergence: How AI Shapes Communication Patterns

Beyond increasing efficiency, AI assistance also influences the way customer service agents communicate. By analyzing changes in text patterns before and after AI adoption, we can understand how AI shapes communication styles over time.

Before the introduction of AI, agents demonstrated a stable and consistent communication style, reflecting both personal habits and the nature of customer interactions. However, after AI adoption, a noticeable shift occurs: agents begin to write differently from their pre-AI baseline. This change is not just a matter of typing faster—the actual content and structure of their messages evolve.

The transformation is particularly pronounced among lower-skill agents, who modify their communication patterns more than their higher-skilled counterparts.

Bridging the Communication Gap Between High- and Low-Skill Agents
A key effect of AI assistance is the increasing similarity in the way high- and low-skill agents communicate. Before AI, the language used by these two groups was notably different, with lower-skilled agents demonstrating more variability in their phrasing and structure. However, once AI is introduced, these differences begin to diminish, leading to a more uniform communication style across agents.

This convergence suggests that AI may serve as a standardizing force, guiding lower-skill agents toward a style of communication that resembles that of their more experienced peers. Since AI-generated recommendations often draw from best practices, agents who follow AI suggestions are likely to adopt clearer, more structured, and more effective messaging over time.

Potential Explanations for Textual Convergence Several factors may contribute to this shift:
AI as a Writing Guide: AI recommendations may introduce new phrasing, tone, or structure that agents integrate into their responses. Over time, this leads to a more consistent and polished communication style. Learning Effects: As lower-skilled agents repeatedly interact with AI-generated suggestions, they may internalize better communication practices, gradually improving their writing skills.

Standardization of Responses: AI often suggests responses based on previously successful customer interactions. This encourages uniformity in phrasing, reducing stylistic differences between agents.
Changes in Customer Interactions: The shift in textual patterns may not only be influenced by AI but also by changes in the nature of customer inquiries. If AI enables agents to handle different types of queries more efficiently, the overall conversation landscape may shift, impacting the way agents communicate.

Implications for AI-Assisted Workflows while AI-driven textual convergence can help maintain consistency and professionalism in customer service interactions, it also raises important considerations Loss of Personalization: If all agents begin to communicate in the same way, customer interactions may become overly standardized, potentially reducing the sense of personalization in support experiences.
Over-Reliance on AI: Agents may default to AI-generated suggestions without critically evaluating them, potentially leading to less nuanced responses in complex situations.
Impact on Creativity and Problem-Solving: If AI encourages uniformity in communication, it may also discourage diverse approaches to problem-solving, particularly in cases where creativity and adaptability are required.
Balancing AI Assistance and Human Judgment

Ultimately, AI-driven textual convergence highlights both the benefits and challenges of AI integration in customer service. While AI helps standardize communication and improve the clarity and effectiveness of responses, organizations must ensure that agents retain the flexibility to personalize interactions and apply their own expertise. By leveraging AI as a supporting tool rather than a rigid template, businesses can strike a balance between efficiency, consistency, and the human touch in customer communication.

5. Effects on the Work Experience

Research indicates that working conditions in contact centers can be challenging. Agents frequently handle emotionally charged conversations, absorbing customer frustrations while maintaining professionalism (Hochschild, 2019). This emotional labor can be a major contributor to stress, burnout, and high turnover in the industry (Lee, 2015). Additionally, many U.S.-based contact centers outsource work to countries like India and the Philippines, requiring agents to work difficult hours and navigate cultural differences, which may sometimes lead to biases from customers.
Moreover, while AI-driven productivity gains might enhance efficiency, they do not necessarily improve workplace satisfaction—especially if workers feel pressured to meet faster response times.

5.1 Customer Sentiment

Customer service agents often face verbal abuse, profanity, and aggressive messages, such as those written in all caps to indicate shouting. AI assistance could influence how customers interact with agents, but the effects may vary. On one hand, AI might lead to smoother interactions by helping agents manage expectations and resolve issues more efficiently. On the other hand, if AI-generated responses come across as overly formal or impersonal, customers might become even more frustrated.
To analyze the impact of AI on customer interactions, we assess sentiment in conversations using natural language processing techniques. Specifically, we employ SiEBERT, a language model trained for sentiment analysis across multiple datasets, including social media and product reviews (Hartmann et al., 2023). Sentiment scores range from -1 (very negative) to 1 (highly positive). We calculate separate sentiment scores for customer and agent responses in each conversation, then aggregate these scores at a monthly level for each agent to track overall trends. Customer sentiment in service interactions varies widely, though, on average, it tends to be mildly positive. However, there are clear outliers—some interactions are exceptionally positive, while others are highly negative. In contrast, agent sentiment remains consistently high, as agents are trained to be courteous and professional in all interactions, even before AI assistance.

The introduction of AI assistance brings a noticeable and lasting improvement in customer sentiment. AI helps agents communicate more effectively, resolve issues faster, and manage customer expectations more clearly. As a result, customers tend to respond more positively, showing a measurable increase in sentiment scores. The improvement is substantial, with customer sentiment increasing by a significant margin, suggesting that AIassisted conversations lead to better overall experiences for customers.

Interestingly, while customer sentiment improves, agent sentiment remains largely unchanged. This is expected, as agents were already trained to maintain a highly positive tone in interactions, and AI assistance does not significantly alter their pre-existing communication style.

AI Benefits Different Agents

The impact of AI assistance is not uniform across all agents. While AI improves customer sentiment across the board, the greatest benefits are seen among agents with lower skill levels or less experience. Lessexperienced agents, who may struggle with handling difficult conversations or demonstrating empathy effectively, gain the most from AI-driven guidance. By providing well-structured, empathetic responses, AI helps these agents navigate customer interactions more smoothly, leading to improved customer treatment. On the other hand, highly skilled and experienced agents see smaller improvements. This aligns with broader findings on AI's productivity impact—experienced agents are already adept at managing customer interactions, so AI offers them fewer additional benefits.

AI Enhances Customer Interactions Several factors may explain why AI improves customer sentiment: Better Communication Strategies: AI-generated responses emphasize clarity, empathy, and professionalism, helping agents communicate in a way that reassures and satisfies customers. Faster Issue Resolution: AI assistance reduces the time needed to resolve common problems, minimizing customer frustration and improving their overall experience. Consistency Across Agents: By guiding less-experienced agents toward more effective communication, AI helps standardize interactions, ensuring customers receive high-quality service regardless of the agent they interact with.

Emotional Influence: When agents handle interactions more smoothly, customers may perceive them as more competent and empathetic, which in turn fosters more positive exchanges. These findings suggest that AI is not just a tool for improving efficiency—it also plays a role in shaping the emotional tone of customer interactions. By guiding agents toward more effective and empathetic communication, AI can create a more positive experience for both customers and workers.

5.2 Customer Confidence and Managerial Escalation

The ability of frontline customer service agents to resolve issues independently plays a crucial role in shaping both customer confidence and organizational efficiency. In traditional service settings, customers occasionally request to escalate their concerns to a supervisor, often due to frustration or a perceived lack of competence from the agent handling their inquiry. These requests for escalation can disrupt workflows, increase managerial workload, and indicate areas where agents may struggle to meet customer expectations.

AI Reduces Escalation Requests

The introduction of AI assistance has a notable impact on the frequency of customers requesting to speak to a manager. By providing agents with AI-powered recommendations, clearer communication strategies, and structured responses, AI enhances their ability to resolve customer issues efficiently and with greater confidence. As a result, customers are less likely to feel the need to escalate their concerns, leading to a measurable decline in such requests.

This reduction suggests that AI not only improves agents' productivity but also increases customer trust in their ability to handle inquiries. When responses are more precise and well-structured, customers are more likely to feel reassured that their concerns are being addressed competently. Over time, this builds greater confidence in frontline agents, reducing the perceived need for managerial intervention.

AI’s impact on escalation requests is particularly pronounced among less-experienced and lower-skilled agents. Before AI adoption, these agents were more likely to struggle with complex inquiries, leading to a higher frequency of escalation requests. However, with AI support, they gain access to structured guidance and suggested responses, enabling them to handle a wider range of customer issues independently. As a result, the decline in escalation requests is more significant for these agents, highlighting AI’s role in bridging skill gaps within the workforce.

For more experienced agents, the impact is less pronounced. Since they already possess strong problemsolving skills and customer-handling expertise, AI serves as a supplementary tool rather than a transformational aid. Nevertheless, even among these agents, AI can help streamline workflows and reduce unnecessary escalations by reinforcing best practices and optimizing communication strategies.

Increased Efficiency: Reducing escalations allows managers and supervisors to focus on more complex issues rather than intervening in routine inquiries, leading to better resource allocation.

Enhanced Customer Satisfaction: When agents can resolve concerns effectively without escalation, customers experience smoother interactions and faster resolutions, contributing to a more positive service experience.

Improved Agent Confidence: Agents who successfully handle issues independently are likely to feel more competent and engaged in their roles, which can contribute to higher job satisfaction and lower turnover rates.

Cost Savings: Fewer escalations mean reduced managerial workload, which can lead to operational cost savings by optimizing staffing needs at higher levels of the organization.
Striking a Balance Between AI Assistance and Human Judgment While AI helps reduce the need for managerial intervention, it is essential to ensure that agents still develop independent problem-solving skills rather than relying solely on AI-generated suggestions. Organizations should encourage a balanced approach where AI serves as a supportive tool rather than a rigid directive. By integrating AI assistance with ongoing training and professional development, businesses can maximize the benefits of AI while maintaining human expertise and adaptability in customer service interactions.

5.3 Attrition: The Impact of AI on Workforce Stability The introduction of generative AI tools in customer service roles can influence various aspects of an agent's work experience, including productivity, stress levels, customer interactions, and overall job satisfaction. While it is challenging to measure these factors directly, examining employee attrition rates provides valuable insights into how AI assistance affects workforce stability.

AI's Role in Reducing Employee Turnover

One of the key findings from our analysis is that AI assistance is associated with a significant reduction in attrition, particularly among newer agents with less than six months of experience. Historically, early-stage attrition has been a major challenge in customer service roles, as new hires often struggle with high-pressure interactions, complex queries, and the emotional toll of handling customer frustrations.
AI tools help mitigate these challenges by providing structured guidance, suggested responses, and enhanced support in real-time. This allows lessexperienced agents to handle customer inquiries with greater confidence and efficiency, reducing frustration and increasing job satisfaction. The result is a notable decrease in turnover rates—approximately 40% lower than the baseline attrition rate for this group.

Attrition Trends Across Different Skill Levels

While AI-driven reductions in attrition are most pronounced among newer employees, AI assistance appears to lower turnover rates across all skill levels. Regardless of experience, agents benefit from AI's ability to streamline workflows, reduce the cognitive burden of crafting responses, and improve overall job performance. However, unlike the clear impact seen among newer employees, the relationship between AI assistance and attrition for more experienced agents does not follow a strict pattern.
This suggests that while AI generally contributes to a more manageable work environment, its effects on longterm employee retention may vary depending on individual experiences, job expectations, and workplace conditions. Some agents may find AI beneficial in reducing their workload, while others might feel constrained by standardized AI-generated suggestions, potentially leading to mixed responses in attrition rates across different skill groups.

6. Conclusion

Advancements in artificial intelligence are creating new economic opportunities, transforming workplace dynamics, and reshaping productivity. This study provides one of the first empirical analyses of the realworld impact of a generative AI tool in a professional setting. Our findings reveal that AI-generated recommendations enhance overall worker productivity by 15%, with the most significant improvements seen among less-experienced and lower-skilled agents. These productivity gains suggest that AI does more than just provide automated suggestions—it contributes to longterm skill development rather than mere reliance on machine-generated guidance. Additionally, AI support enhances the workplace experience by improving customer sentiment, boosting agent confidence, and reducing employee turnover.

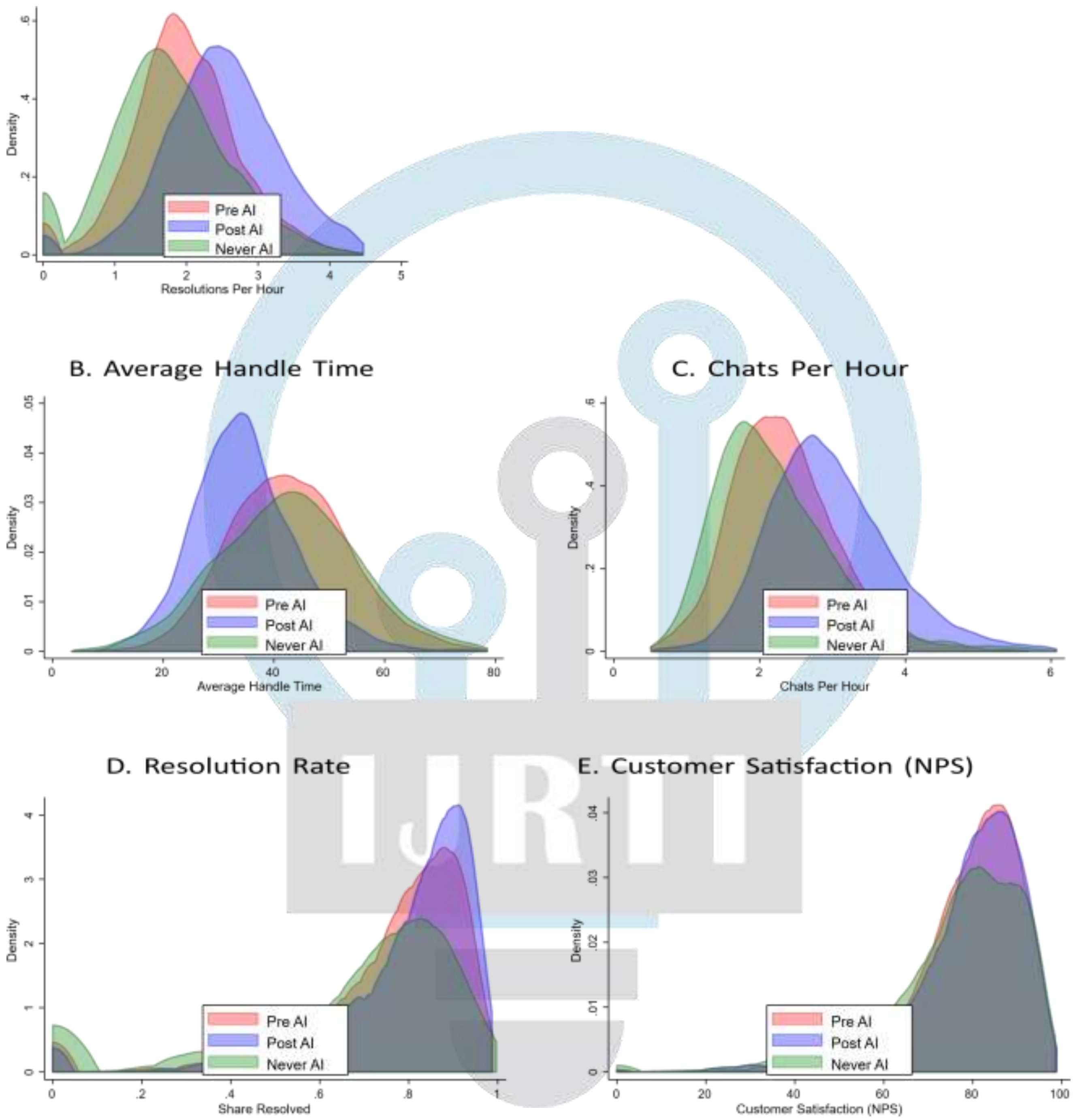
Limitations and Open Questions

While our analysis provides valuable insights, it is important to acknowledge certain limitations. First, our findings are based on a single AI tool implemented within one company and occupation. This specific work environment involves a relatively stable set of products and customer inquiries, which may not be representative of other industries where rapid changes require constant adaptation. In such cases, AI-generated recommendations might either help by synthesizing evolving best practices or hinder progress by reinforcing outdated methods from historical training data. Prior studies (Otis et al., 2023; Perry et al., 2022) have highlighted instances where AI adoption has led to mixed or even negative outcomes.
Second, our study focuses on short- to medium-term effects without capturing broader market dynamics. While we lack direct access to compensation data, conversations with managers suggest that AI assistance may have led to higher performance-based pay, as bonuses were linked to metrics such as resolution rates and average handling time. However, such financial incentives may not be sustainable in the long run, as companies frequently adjust performance benchmarks once too many employees begin meeting targets. This phenomenon, often referred to as the "ratchet effect," could offset some of the initial benefits of AI adoption. Another crucial aspect is the long-term impact of AI on workforce demand and job structures. If customer inquiries remain relatively constant, AI-driven efficiency gains could reduce the overall need for human labor. A preliminary estimate suggests that the company could handle the same number of customer support cases with 12% fewer working hours. Conversely, improvements in service quality and efficiency could encourage more customers to seek support, thereby increasing demand for customer service jobs (Berg et al., 2018; Korinek, 2022). Moreover, AI adoption could lead to the creation of new roles, such as monitoring and refining AIgenerated responses. Some companies have already begun assigning experienced agents to oversee AI recommendations and refine training models (Autor et al., 2022).
Additionally, even minimal AI adoption can influence market trends, wages, and productivity at a broader scale (Raymond, 2023). The full impact of AI on labor markets will depend on factors such as industry-specific needs, evolving workplace policies, and

access to AI training programs (Brynjolfsson et al., 2021). Research has shown that employees' willingness to engage with AI tools depends heavily on adequate training and compliance with regulations governing data privacy and confidentiality (Humlum & Vestergaard, 2024).

Redefining Worker Productivity in the AI Era The findings also raise critical questions about how worker productivity is defined in AI-integrated environments. Traditionally, customer service agent productivity has been measured by their ability to assist individual customers. However, in a setting where AI models are continuously trained on real-world interactions, an agent's contributions extend beyond individual problem-solving. Their responses provide data that refine AI models, shaping future interactions for other agents. High-performing employees play a key role in generating the training data that strengthens AI recommendations, yet AI assistance could unintentionally reduce their incentive to innovate and develop new solutions. Moving forward, compensation strategies that encourage workers to contribute to AI model improvement may become essential.
As generative AI continues to evolve, further research is needed to explore these emerging questions.
Understanding the long-term effects of AI on job design, worker incentives, and labor market equilibrium will be crucial in shaping policies that maximize the benefits of AI while mitigating potential risks.

Figure 1: Raw Productivity Distributions, by AI Treatment A. Resolutions Per Hour



A.3 Empirical Specifications

A.3.1 Pre-treatment Worker Skill Specification

$$y_{it} = \delta_t + \alpha_i + \sum_{q=1}^{5} \beta_q (AI_{it} \times Q_{iq}) + \gamma X_{it} + \epsilon_{it} \qquad (2)$$

Our worker skill specification allows us to estimate how the impact of assistance varies by agent skill when they receive access to AI assistance. The regression is conducted at the agent-month level, where:

- y_{it} is the resolutions per hour of agent i in year-month t.
- δ_t represents year-month fixed effects.

- α_i denotes agent-level fixed effects.

- AI_{it} is an indicator equal to 1 if agent *i* has access to AI assistance at time *t*, 0 otherwise.

- Q_{iq} is an indicator function that equals 1 if agent *i* belonged to skill quintile *q* at the time of treatment, where *q* ranges from 1 (lowest skill) to 5 (highest skill).

- X_{it} is a set of time-varying controls, specifically fixed effects for agent tenure in months.

- β_q represents the average treatment effect of AI assistance for agents in skill quintile *q*

We estimate this equation separately for each of our outcome variables, which include our main measure of productivity for agent *i* in year-month *t* (resolutions per hour), as well as call resolution rate, customer satisfaction score, average call duration, calls handled per hour and requests to speak to the manager. We cluster standard errors at the agent level to account for within-agent correlations in the error terms. In our main analysis, we find very similar results across estimators and similar main effects across adoption cohorts, so we estimate this regression specification using
OLS.

A.3.2 Pre-treatment Worker Tenure Specification

5

$$y_{it} “ \delta_t ` \alpha_i ` ÿ \beta_e p AI_{it} ˆ Exp_{ie} q ` \epsilon_{it} \quad (3)$$

e“1

Our tenure specification allows us to estimate how the impact of assistance varies by agent experience when they receive access to AI assistance. The regression is conducted at the agent- month level, where:

- y_{it} is the resolutions per hour of agent *i* in year-month *t*.

- δ_t represents year-month fixed effects.

- α_i denotes agent-level fixed effects.

- AI_{it} is an indicator equal to 1 if agent *i* has access to AI assistance at time *t*, 0 otherwise.

- Exp_{ie} is an indicator that equals 1 if agent *i* has *e* months of experience at the time of treatment, and 0 otherwise. We divide months of experience into five categories: agents in their first month on the job (“0 months”), 1-2 months of experience, 3-6 months of experience, 7-12 months, and over 12 months of experience.

- X_i includes a set of fixed effects for agent *i*’s skill quintile at the time of treatment. This is only available for treated agents and is time-invariant.

- β_e is the average treatment effect of AI assistance for agents with *e* months of experience when treated.

We estimate this equation separately for each of our outcome variables, which include our main measure of productivity for agent *i* in year-month *t* (resolutions per hour), as well as call resolution rate, customer satisfaction score, average call duration, and calls handled per hour. The skill quintile at AI treatment is time invariant and is not defined for control group workers. Standard errors are clustered at the agent level, and we estimate this regression using OLS.

A.3.3 Adherence to AI recommendations

5

$$y_{it} “ \delta_t ` \alpha_i ` ÿ \beta_a p AI_{it} ˆ Adh_{ia} q ` \gamma X_{it} ` \epsilon_{it} \quad (4)$$

$a=1$

This specification allows us to estimate how the impact of AI assistance varies by treated agents' adherence in their first month of AI access. The regression is conducted at the agent-month level, where:

- y_{it} is the resolutions per hour of agent i in year-month t.
- δ_t represents year-month fixed effects.
- α_i denotes agent-level fixed effects.
- AI_{it} is an indicator equal to 1 if agent i has access to AI assistance at time t, 0 otherwise.
- Adh_{ia} is an indicator equal to 1 if agent i is in the ath quintile of adherence in their first month of access, 0 otherwise.
- X_{it} includes time-varying controls, specifically fixed effects for agent tenure.
- β_a is the average impact of AI assistance for agents in the ath quintile of initial adherence.

Standard errors are clustered at the agent level. We estimate this regression with OLS.

A.3.4 Heterogeneity by Chat Topic

$$y_{itc} = \delta_t + \alpha_i + \sum_{r=1}^{4} \beta_r (AI_{it} \times Topic_{cr}) + \gamma X_{itc} + \epsilon_{itc} \tag{5}$$

This topic-based specification allows us to estimate how the impact of AI assistance varies by the routine nature of customers' problems. The regression is conducted at the chat level, where:

- y_{itc} is the duration of chat c assigned to agent i in year-month t.
- δ_t represents year-month fixed effects.
- α_i denotes agent-level fixed effects.
- AI_{it} is an indicator equal to 1 if agent i has access to AI assistance at time t, 0 otherwise.
- $Topic_{cr}$ is an indicator equal to 1 if chat c belongs to topic frequency category r (where r ranges from 1 to 4), 0 otherwise. The frequency of the topic is defined using frequency across all chats and agents.
- X_{it} includes fixed effects for agent tenure and the overall category of topic frequency ($Topic_{cr}$).
- β_r estimates the impact of AI assistance on chat duration for chats in the topic frequency category r.

Standard errors are clustered at the agent level and we estimate this regression with OLS. We drop the small number of topics that are classified as unsure or "other." We also estimate a separate specification on the agent-specific frequency of the technical support issue.

$$y_{itc} = \delta_t + \alpha_i + \sum_{r=1}^{4} \beta_r (AI_{it} \times AgentTopic_{icr}) + \gamma X_{cit} + \epsilon_{itc} \tag{6}$$

- $AgentTopic_{icr}$ is an indicator equal to 1 if chat c belongs to agent-specific topic frequency category r (where r ranges from 1 to 4), 0 otherwise. Topic frequency is defined relative to conversations conducted by agent i.
- X_{cit} includes time-varying controls, specifically fixed effects for agent tenure, aggregate topic category defined over all conversations ($Topic_{cr}$), and agent-specific topic frequency category ($AgentTopic_{icr}$).

– β_r estimates the impact of AI assistance on chat duration for chats in agent-specific topic frequency category r.

A.3.5 Attrition

$$attrit_{it} “ \delta_t ` \beta AI_{it} ` \gamma X_{it} ` \epsilon_{it} \tag{7}$$

This specification allows us to look at how AI assistance impacts agent attrition. The regression is conducted at the agent-month level, where:

– $attrit_{it}$ is equal to 1 if agent i leaves in year-month t.

– δ_t represents year-month fixed effects.

– AI_{it} is an indicator equal to 1 if agent i has access to AI assistance at time t, 0 otherwise.

– X_{it} includes time-varying controls, specifically fixed effects for agent tenure, agent location, country and company employing the agent (data firm or subcontractor).

– β_a is the average impact of AI assistance on attrition.

Attrition captures both voluntary or involuntary separations, which are undistinguishable in our data. Standard errors are clustered at the agent level and we estimate this regression with OLS. We cannot include for agent-fixed effects, because agents only leave once, so agent-fixed effects are colinear with attrition. We drop all observations for treated agents before treatment because, by construction, agents must survive through treatment to receive AI assistance. We also estimate specifications where we estimate the attrition effects by agent tenure and skill at AI deployment.

II. Acknowledgment

I would like to express my sincere gratitude to **Prof. Shri Krishna Balwante** for his invaluable guidance, encouragement, and support throughout the course of this research. His insights and expertise have been instrumental in shaping this paper and enhancing the depth of my understanding. I am truly thankful for his mentorship and continuous inspiration.

References

[1] **Alan Akbik, Duncan Blythe, and Roland Vollgraf.** 2018. Contextual string embeddings for sequence labeling. In Proceedings of the 27th International
[2] Conference on Computational Linguistics, pages 1638–1649.
[3] **Rami Al-Rfou, Dokook Choe, Noah Constant, Mandy Guo, and Llion Jones**. 2018. Character-level language modeling with deeper self-attention. arXiv preprint arXiv:1808.04444.
[4] **Rie Kubota Ando and Tong Zhang**. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. Journal of Machine Learning Research, 6(Nov):1817–1853.
[5] **Luisa Bentivogli, Bernardo Magnini, Ido Dagan,**
[6] **Hoa Trang Dang, and Danilo Giampiccolo.** 2009. The fifth PASCAL recognizing textual entailment challenge. In TAC. NIST.
[7] **John Blitzer, Ryan McDonald, and Fernando Pereira**. 2006. Domain adaptation with structural correspondence learning. In Proceedings of the 2006 conference on empirical methods in natural language processing, pages 120–128. Association for Computational Linguistics. **Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher** D. Manning. 2015. A large annotated corpus for learning natural language inference.
[8] In EMNLP. Association for Computational Linguistics.
[9] **Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai**. 1992.
[10] Class-based n-gram models of natural language.
[11] Computational linguistics, 18(4):467–479.
[12] **Daniel Cer, Mona Diab, Eneko Agirre, Inigo**

[13] **LopezGazpio, and Lucia Specia**. 2017. Semeval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 1–14, Vancouver, Canada. Association for Computational Linguistics.
[14] **Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge,**
[15] **Thorsten Brants, Phillipp Koehn, and Tony Robinson**. 2013. One billion word benchmark for measuring progress in statistical language modeling. arXiv preprint arXiv:1312.3005.
[16] **Z. Chen, H. Zhang, X. Zhang, and L. Zhao**. 2018.
[17] Quora question pairs.
[18] Christopher Clark and Matt Gardner. 2018. Simple and effective multi-paragraph reading comprehension. In ACL.
[19] **Kevin Clark, Minh-Thang Luong, Christopher D Manning, and Quoc Le**. 2018. Semi-supervised sequence modeling with cross-view training. In Proceedings of the 2018 Conference on Empirical Methods in Natural
[20] Language Processing, pages 1914– 1925.
[21] **Ronan Collobert and Jason Weston**. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on
[22] Machine learning, pages 160–167. ACM.
[23] **Alexis Conneau, Douwe Kiela, Holger Schwenk, Lo¨ıc Barrault, and Antoine Bordes**. 2017. Supervised learning of universal sentence representations from natural language inference data. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 670–680, Copenhagen,
[24] Denmark. Association for Computational
[25] Linguistics.
[26] **Borusyak, Kirill, Xavier Jaravel, and Jann Spiess,** "Revisiting Event Study Designs: Robust and Efficient Estimation," 2022.
[27] **Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M**. Hitt, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," The Quarterly Journal of Economics, 02 2002, 117 (1), 339–376.
[28] **Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz**
[29] **Litwin, Scott Gray, Benjamin Chess, Jack Clark,**
[30] **Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei**, "Language Models are Few-Shot Learners," July 2020. arXiv:2005.14165 [cs].
[31] Brynjolfsson, Erik and Tom Mitchell, "What Can Machine Learning, Do? Workforce Implications," Science, December 2017, 358, 1530–1534.
[32] **Daniel Rock, and Chad Syverson**, "The Productivity JCurve: How Intangibles Complement General Purpose Technologies," American Economic Journal:
[33] Macroeconomics, January 2021, 13 (1), 333–72.
[34] **Bubeck, Sebastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin** Tat Lee, Yuanzhi Li, Scott Lundberg et al., "Sparks of artificial general intelligence: Early experiments with gpt-4," arXiv preprint arXiv:2303.12712, 2023.
[35] **Buesing, Eric, Vinay Gupta, Sarah Higgins, and Raelyn Jacobson**, "Customer care: The future talent factory," Technical Report, McKinsey & Company June 2020. Callaway, Brantly and Pedro H. C. Sant'Anna, "Differencein-Differences with multiple time periods," Journal of Econometrics, December 2021, 225 (2), 200–230.
[36] **Calvino, Flavio and Luca Fontanelli,** "A Portrait of AI Adopters across Countries: Firm Characteristics, Assets' Complementarities and Productivity," Technical Report, OECD, Paris April 2023.
[37] **Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer**, "The Effect of Minimum Wages on Low-Wage Jobs*," The Quarterly Journal of Economics, May 2019, 134 (3), 1405–1454.
[38] **Choi, Jonathan H. and Daniel Schwarcz**, "AI Assistance in Legal Analysis: An Empirical Study," August 2023.
[39] Choi, Jun Ho, Oliver Garrod, Paul Atherton, Andrew JoyceGibbons, Miriam MasonSesay, and Daniel Björkegren, "Are LLMs Useful in the Poorest Schools? TheTeacher.AI in Sierra Leone," 2024.
[40] **Chui, Michael, Bryce Hall, Alex Singla, and Alex Sukharevsky**, "Global survey: The state of AI in 2021," Technical Report, McKinsey & Company 2021.
[41] de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," American Economic Review, September 2020, 110 (9), 2964–96.
[42] **Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Lifshitz-Assaf, Katherine**
[43] **Kellogg, Saran Rajendran, Lisa Krayer, Francois Candelon, and Karim Lakhani**, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," September 2023.
[44] **Dixon, Jay, Bryan Hong, and Lynn Wu,** "The Robot Revolution: Managerial and Employment Consequences for Firms," DecisionSciRN: Other Performance Management (Sub-Topic), 2020.

[45] **Dunn, Andrew, Diana Inkpen, and Răzvan Andonie,** "Context-Sensitive Visualization of Deep Learning Natural Language Processing Models," 2021.

[46] Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," March 2023. arXiv:2303.10130 [cs, econ, q-fin].

[47] **Felten, Edward W., Manav Raj, and Robert Seamans**, "Occupational Heterogeneity in Exposure to Generative AI," April 2023.

[48] **Garicano, Luis**, "Hierarchies and the Organization of Knowledge in Production," Journal of Political Economy, 2000, 108 (5), 874–904. Publisher: The University of Chicago Press.

[49] and Esteban Rossi-Hansberg, "Knowledge-Based

[50] Hierarchies: Using Organizations to Un- derstand the Economy," Annual Review of Economics, 2015, 7 (1), 1–30.

[51] **Gemini Team, "Gemini**: A Family of Highly Capable Multimodal Models," Working Papers, Google 2024.

[52] Goldin, Claudia and Lawrence F. Katz, "The Origins of Technology-Skill Complementarity*," The Quarterly Journal of Economics, 08 1998, 113 (3), 693–732. and , The Race between Education and Technology, Harvard University Press, 2008.

[53] Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," Journal of Econometrics, December 2021, 225 (2), 254–277.

[54] **Google, "AI vs. Machine Learning:** How Do They Differ?" **Gretz, Whitney and Raelyn Jacobson**, "Boosting contactcenter performance through employee engagement," Technical Report, McKinsey & Company 2018.

[55] Halevy, Alon, Peter Norvig, and Fernando Pereira, "The Unreasonable Effectiveness of Data," IEEE Intelligent Systems, March 2009, 24 (2), 8–12.

[56] **Hartmann, Jochen, Mark Heitmann, Christian Siebert, and Christina Schamp**, "More than a Feeling: Accuracy and Application of Sentiment Analysis," International Journal of Research in Marketing, 2023, 40 (1), 75–87.

[57] Hochschild, Arlie Russell, The managed heart: Commercialization of human feeling, University of California press, 2019.

[58] **Hoffman, Mitchell, Lisa B Kahn, and Danielle Li, "**Discretion in Hiring*," The Quarterly Journal of Economics, 10 2017, 133 (2), 765–800.

[59] Hugging Face, "sentence-transformers/all-MiniLM-L6v2," April 2023.

[60] **Humlum, Anders and Emilie Vestergaard**, "The Adoption of ChatGPT," May 2024.

[61] **Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe**, "AI, Skill, and Productivity: The Case of Taxi Drivers," Working Paper 30612, National Bureau of Economic Research October 2022.

[62] **Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei,** "Scaling laws for neural language models," arXiv preprint arXiv:2001.08361, 2020.

[63] **Katz, Lawrence F. and Kevin M. Murphy,** "Changes in Relative Wages, 1963-1987: Supply and Demand

[64] Factors," The Quarterly Journal of Economics, 1992, 107 (1), 35–78.

[65] **Korinek, Anton,** "How innovation affects labor markets: An impact assessment," Working Paper, Brookings

[66] Institution June 2022.

[67] **Koroteev, M. V., "BERT**: A Review of Applications in Natural Language Processing and Understanding," 2021.

[68] Kumar, Harsh, David Rothschild, Daniel Goldstein, and Jake Hofman, "Math Education with Large Language Models: Peril or Promise?," SSRN Electronic Journal, 01 2023.

[69] **Lee, Don**, "The Philippines has become the call-center capital of the world," Los Angeles Times, February 2015. Section: Business.

[70] **Legg, Shane, Marcus Hutter et al.,** "A collection of definitions of intelligence," Frontiers in Artificial Intelligence and applications, 2007, 157, 17. **Li, Chun**, "OpenAI's GPT-3 Language Model: A Technical Overview," June 2020.

[71] **Liu, Yiheng, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, Zihao Wu, Dajiang Zhu, Xiang Li, Ning Qiang, Dingang Shen, Tianming Liu, and Bao Ge**, "Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models," April 2023. arXiv:2304.01852 [cs].

[72] **Meijer, Erik,** "Behind every great deep learning framework is an even greater programming languages concept (keynote)," in "Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering

[73] Conference and Symposium on the Foundations of Software Engineering" 2018, pp. 1–1.

[74] **Mejova, Yelena, "Sentiment Analysis**: An Overview," University of Iowa, Computer Science Department, 2009.

[75] Michaels, Guy, Ashwini Natraj, and John Van Reenen,"Has ICT Polarized Skill Demand? Evidence from Eleven

[76] Countries Over Twenty-Five Years," The Review of Economics and Statistics, 2014, 96 (1), 60–77.

[77] **Nguyen, Nhan and Sarah Nadi**, "An Empirical Evaluation of GitHub Copilot's Code Suggestions," in "2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR)" May 2022, pp. 1–5. ISSN:2574-3864.

[78] **Noy, Shakked and Whitney Zhang**, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," Available at SSRN 4375283, 2023.

[79] OECD, OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market, Paris: Organisation for Economic Co-operation and Development, 2023.