

e-ISSN:2582-7219



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 6, Issue 3, March 2023



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.54



6381 907 438



6381 907 438



ijmrset@gmail.com



www.ijmrset.com



# Advancements and Applications of Generative Artificial Intelligence and show the Experimental Evidence on the Productivity Effects using Generative Artificial Intelligence

Sakshi Sharma<sup>1</sup>, Natasha Dutta<sup>2</sup>

Senior Technical Project Manager, Kforce Global Solutions Inc. <sup>1</sup>

Information Security Engineer, Online Micro Services, India<sup>2</sup>

**ABSTRACT:** We investigate the productivity impacts of a generative artificial intelligence technology—specifically, the assistive chatbot ChatGPT—within the realm of mid-level professional writing tasks. In a preregistered online experiment, we assigned occupation-specific, incentivized writing tasks to 444 college-educated professionals, with half of the participants randomly exposed to ChatGPT. Our findings reveal that ChatGPT significantly enhances average productivity: the time taken to complete tasks decreases by 0.8 standard deviations, and output quality improves by 0.4 standard deviations. Additionally, the use of ChatGPT reduces inequality among workers by compressing the productivity distribution, primarily benefiting those with lower initial ability. ChatGPT primarily substitutes for worker effort rather than complementing worker skills, shifting the focus of tasks from rough-drafting to idea generation and editing. Furthermore, exposure to ChatGPT increases job satisfaction and self-efficacy while also heightening both concerns and excitement about automation technologies.

**KEYWORDS:** ChatGPT, AI, GAI, Technology, IT

## I. INTRODUCTION

The advent of generative artificial intelligence (GAI) technologies has marked a significant milestone in the evolution of AI, offering transformative potential across various professional domains. One notable application of GAI is the assistive chatbot, ChatGPT, which has shown promise in enhancing productivity and efficiency in mid-level professional writing tasks. As organizations and individuals increasingly rely on AI tools to streamline their workflows, understanding the specific impacts of such technologies on productivity is crucial. This study focuses on evaluating the productivity effects of ChatGPT in professional writing contexts. We conducted a preregistered online experiment involving 444 college-educated professionals who were assigned occupation-specific, incentivized writing tasks. Half of these participants were randomly selected to use ChatGPT during their tasks. Our objective was to measure changes in productivity, time efficiency, and output quality attributable to the use of ChatGPT. The results of our experiment provide compelling evidence that ChatGPT significantly enhances average productivity by reducing the time required to complete tasks and improving the quality of the output. Notably, the technology appears to benefit workers with lower initial abilities more, thereby reducing productivity inequality. The introduction of ChatGPT primarily substitutes for manual effort, shifting the nature of tasks from rough drafting to more creative and refined processes such as idea generation and editing. Furthermore, our findings indicate that exposure to ChatGPT boosts job satisfaction and self-efficacy among users, while also eliciting heightened awareness and mixed feelings about the broader implications of automation technologies. This study contributes to the growing body of evidence on the productivity effects of GAI and underscores the importance of evaluating both the benefits and challenges associated with integrating these technologies into professional environments.

This paper pioneers the exploration of the productivity effects of generative artificial intelligence, specifically through the use of ChatGPT, in a professional writing context. To achieve this, we conducted an online experiment involving 444 experienced, college-educated professionals, each tasked with completing two occupation-specific, incentivized writing assignments. These occupations included marketers, grant writers, consultants, data analysts, human resource professionals, and managers, and the tasks ranged from writing press releases and short reports to crafting analysis plans and delicate emails. These assignments, designed to be 20 to 30 minutes in duration, closely mimic real-world tasks typical in these professions, a fact corroborated by our participants, who reported frequent engagement with similar tasks and rated them as realistic representations of their daily work. To motivate high-quality output, participants were offered substantial bonus payments, and their work was assessed by blind evaluations from experienced professionals within the same fields, who were incentivized to grade meticulously. The evaluation criteria



included overall grade, writing quality, content quality, and originality, with each output being reviewed by three evaluators, resulting in an average within-essay cross-evaluator correlation of 0.44.

Half of our participants were randomly assigned to the treatment group and instructed to sign up for ChatGPT between the first and second task. They were guided on its use and informed that they could utilize it for the second task if they found it beneficial. The control group, on the other hand, was directed to sign up for the LaTeX editor Overleaf. This experimental design allowed us to estimate the causal effects of ChatGPT by leveraging both within-person and between-person variations, using performance on the first task as a baseline measure to facilitate our inequality analyses. We collected data on participants' output, total time taken, time allocated to various subtasks, job satisfaction, self-efficacy, and beliefs about automation. We captured snapshots of each participant's output every minute during the task to construct an objective measure of active time on task and to detect any ChatGPT usage within the control group and during the pre-treatment task.

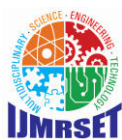
A comprehensive description of our experimental design, along with the survey questionnaires and supplementary figures validating our central measures and extending our main results, is provided in the Online Appendix. Descriptive statistics about our sample, as well as balance and selective attrition tests, are presented in Table 1. We observed an attrition rate of 5% in the control group and 10% in the treatment group. Balance tests across 13 pre-treatment characteristics revealed only minor significant differences in employment status and HR professional representation between the treatment and control groups. However, our partially within-person design, which accounts for pre-treatment task performance, mitigates the influence of selective attrition on our results. Further robustness checks, including Lee (2009) bounds and results controlling for employment status and occupation, confirm the resilience of our findings to selective attrition.

## II. LITERATURE REVIEW

The latest model from OpenAI, GPT-4, was developed using an unprecedented scale of computational power and data. This paper reports on our examination of an early iteration of GPT-4, during its active development phase at OpenAI. We argue that this early version of GPT-4 represents a new class of LLMs, alongside models like ChatGPT and Google's PaLM, which demonstrate a level of general intelligence surpassing that of previous AI systems. We explore the advancing capabilities and implications of these models, showing that GPT-4 not only excels in language mastery but also tackles complex and novel tasks across mathematics, coding, vision, medicine, law, and psychology without requiring specialized prompting. Remarkably, GPT-4's performance closely approaches, and often exceeds, human-level proficiency, outstripping previous models such as ChatGPT. Given the breadth and depth of GPT-4's abilities, it can be considered an early, albeit incomplete, version of an artificial general intelligence (AGI) system. Our analysis focuses on identifying GPT-4's limitations and the challenges involved in developing more advanced AGI systems, which may necessitate moving beyond the paradigm of next-word prediction. We conclude with reflections on the societal impact of this technological advancement and suggest directions for future research.

In healthcare, it could advance diagnostics and improve accessibility, yet it risks exacerbating existing disparities. Each section of the article addresses a specific topic, critically evaluates current research, identifies key gaps, and suggests future research directions, including explicit trade-offs that complicate the formation of a priori hypotheses. We conclude with an analysis of the role of policymaking in maximizing the potential of generative AI to reduce inequalities while mitigating its adverse effects. We assess the strengths and weaknesses of current policy frameworks in the European Union, the United States, and the United Kingdom, noting that none fully addresses the socioeconomic challenges identified. We propose several concrete policies to foster shared prosperity through the advancement of generative AI. This article underscores the necessity for interdisciplinary collaboration to comprehend and tackle the multifaceted challenges posed by generative AI.

These tools offer significant potential benefits to organizations, society, and individuals, but they also pose ethical and legal issues. This article consolidates insights from 43 experts across diverse fields, including computer science, marketing, information systems, education, policy, hospitality, management, publishing, and nursing. The contributors recognize ChatGPT's ability to boost productivity, particularly noting its potential for significant gains in industries such as banking, hospitality and tourism, and information technology, as well as its positive impact on business activities like management and marketing. However, they also address its limitations, including potential disruptions to established practices, privacy and security risks, and issues related to biases, misuse, and misinformation. Opinions are divided on whether the use of ChatGPT should be restricted or regulated. The article highlights key areas requiring further research, including: understanding the skills and resources needed to manage generative AI; investigating biases



in generative AI linked to training datasets; exploring the most suitable contexts for implementing generative AI; determining the best ways to combine human input with generative AI; assessing the accuracy of AI-generated text; and addressing ethical and legal concerns associated with generative AI.

### III. METHODOLOGY

Our study employed a robust experimental design to investigate the productivity impacts of ChatGPT on mid-level professional writing tasks. We recruited 444 college-educated professionals from various occupations, including marketers, grant writers, consultants, data analysts, human resource professionals, and managers, ensuring a diverse and representative sample. Participants were randomly assigned to complete two occupation-specific writing tasks, each designed to closely resemble real-world assignments typical of their professions. These tasks, which included writing press releases, short reports, analysis plans, and delicate emails, were structured to take 20 to 30 minutes, reflecting the time constraints and pressures of actual workplace scenarios. To motivate participants to produce high-quality work, we offered substantial monetary incentives in the form of large bonus payments, aligning their efforts with real-world stakes.

The assessment of task output was conducted by experienced professionals in the same fields, who evaluated the submissions blind to the experimental conditions. These evaluators, incentivized to provide careful and accurate assessments, graded the outputs on overall quality, writing quality, content quality, and originality, with each piece of work reviewed by three different evaluators. The average within-essay cross-evaluator correlation was 0.44, indicating a reasonable level of consistency among evaluators.

For the intervention, 50% of the participants were randomly assigned to the treatment group. These individuals were instructed to sign up for ChatGPT between their first and second tasks. They were given a walkthrough on how to use the tool and informed that they were permitted to use it for the second task if they found it beneficial. The control group, meanwhile, was directed to sign up for the LaTeX editor Overleaf, ensuring that both groups engaged with new software but only the treatment group used generative AI. This approach allowed us to estimate the causal effects of ChatGPT by leveraging both within-person (comparing each participant's performance on the two tasks) and between-person (comparing the treatment group to the control group) variations.

To further enrich our analysis, we collected detailed data on various dimensions of the participants' performance. This included the total time taken to complete the tasks, the time spent on different subcomponents of the tasks, job satisfaction, self-efficacy, and their beliefs about automation. To ensure accurate measurement of task engagement, we took minute-by-minute snapshots of each participant's work, enabling us to construct an objective measure of active time on task and to detect any ChatGPT usage within the control group or on the pre-treatment task.

Our experimental design also incorporated rigorous checks and balances to ensure the reliability and validity of our findings. Descriptive statistics about our sample, as well as balance and selective attrition tests, were carefully documented and analyzed. The attrition rate was 5% in the control group and 10% in the treatment group, which we addressed through balance tests across 13 pre-treatment characteristics. These tests revealed small significant differences in only two characteristics (employment status and being an HR professional) between the treatment and control groups. However, our partly within-person design, which controls for performance on the pre-treatment task, is robust against the potential biases introduced by selective attrition. We further validated our results by reporting Lee (2009) bounds and conducting additional analyses controlling for employment status and occupation, all of which confirmed the robustness of our main findings. This methodology ensures that our findings are both reliable and applicable to real-world professional settings, offering valuable insights into the potential benefits and implications of integrating generative AI technologies into the workplace.

### IV. RESULT

In the treatment group, 92% of participants successfully signed up for ChatGPT, and 81% chose to use it for the second task, rating its usefulness at an average of 4.4 out of 5. Before the treatment, approximately 70% of participants had heard of ChatGPT, and 30% had previously used it. Both self-reported and objective measures indicate that only 10-20% of the control group used ChatGPT during the tasks. This results in at least a 60 percentage point experimentally-induced gap in usage between the treatment and control groups for the second task. The usage of ChatGPT by some control group participants means that our estimates provide conservative lower bounds on the productivity effects of ChatGPT usage.



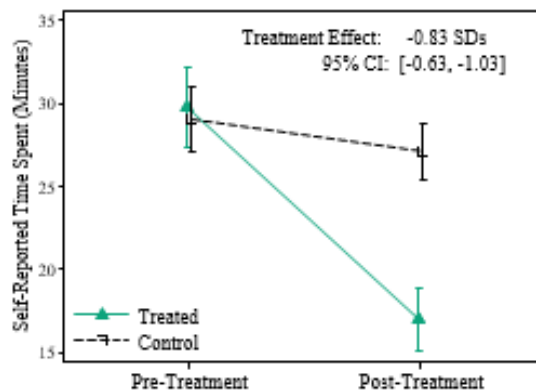
We measure productivity as earnings per minute, and Figure 1 illustrates the significant impact of our experimental intervention on this outcome. In the treatment group, the time taken to complete the post-treatment task decreases by 10 minutes (37%) compared to the control group, which averages 27 minutes ( $p = 0.000$ ). Additionally, the average evaluator grades in the treatment group increase by 0.45 standard deviations ( $p = 0.000$ ), with comparable improvements in overall grades and specific grades for writing quality, content quality, and originality.

Panels (c) and (d) of Figure 1 reveal that these effects are pervasive across the entire time and grade distributions: the entire time distribution shifts to the left, indicating faster task completion, while the entire grade distribution shifts to the right, indicating higher quality work. At the individual worker level, Figure 2 demonstrates that treated workers who received a low grade on the first task see improvements in both grades and time spent on the second task, whereas those who initially received high grades maintain their grade levels while significantly reducing their time spent.

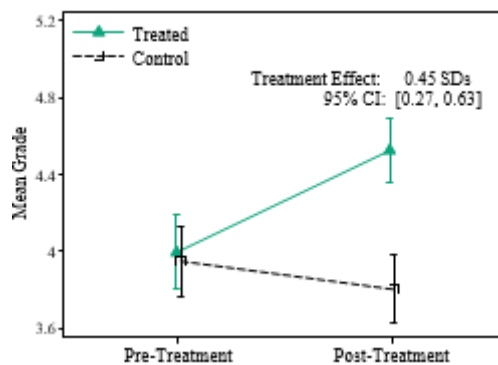
These findings are consistent across our two main incentive schemes, which cover 80% of respondents. In the ‘linear’ scheme, participants are paid \$1 for each point they receive. This consistency underscores the robustness of our results, highlighting the substantial productivity gains attributable to the use of ChatGPT.

In the ‘linear’ scheme, participants are paid \$1 for each submission point they receive, with each submission graded on a 1-7 point scale. In the “convex” scheme, respondents receive an additional \$3 for earning a grade of 6 or 7, providing extra motivation to produce high-quality work.

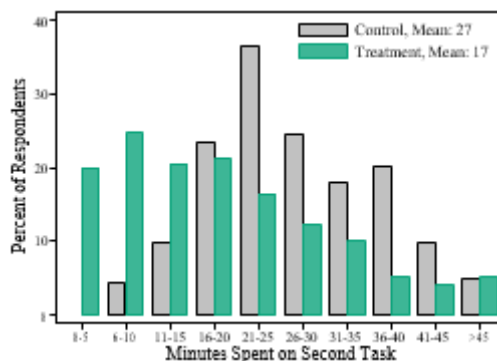
(a) Time Taken Decreases



(b) Average Grade Increase



(c) Time Distribution (second task)



(d) Grades (second task)

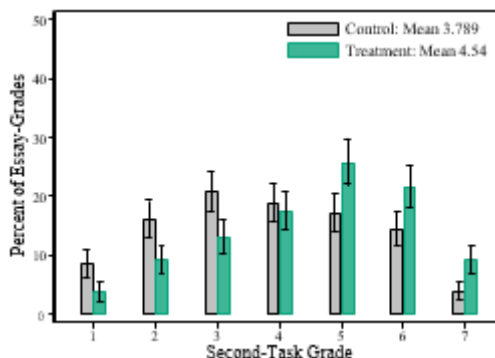


Figure 1: Treatment Effects on Productivity

Two supplementary interventions enable us to delve deeper into the effects of ChatGPT. In one arm, involving 20% of participants, we required both the treatment and control groups to spend exactly 15 minutes on each task. This constraint fixed the effort across groups, allowing us to attribute any difference in grades solely to the impact of ChatGPT access on productive capacity. In this scenario, the treatment led to an increase in grades by 0.39 standard deviations ( $p = 0.13$ ), although this result was imprecisely estimated and there was a slight imbalance in pre-treatment outcomes.

In another arm, involving 30% of the treatment group, participants were shown their output from the first task after completing the second task. They were then given the option to edit or replace their initial response using ChatGPT. Of these participants, 23% chose to replace their original response with ChatGPT’s output, and 25% used ChatGPT to edit their initial response. This behavior indicates that participants view ChatGPT not only as a tool to save time but also as a means to enhance the quality of their work.

Task Structure

As indicated by our findings, ChatGPT significantly alters the structure of writing tasks. Figure 2, Panel A, reveals that before the treatment, participants allocate approximately 25% of their time to brainstorming, 50% to writing a rough draft, and 25% to editing. After the treatment, the proportion of time spent on writing a rough draft decreases by more than half, while the time spent on editing more than doubles.

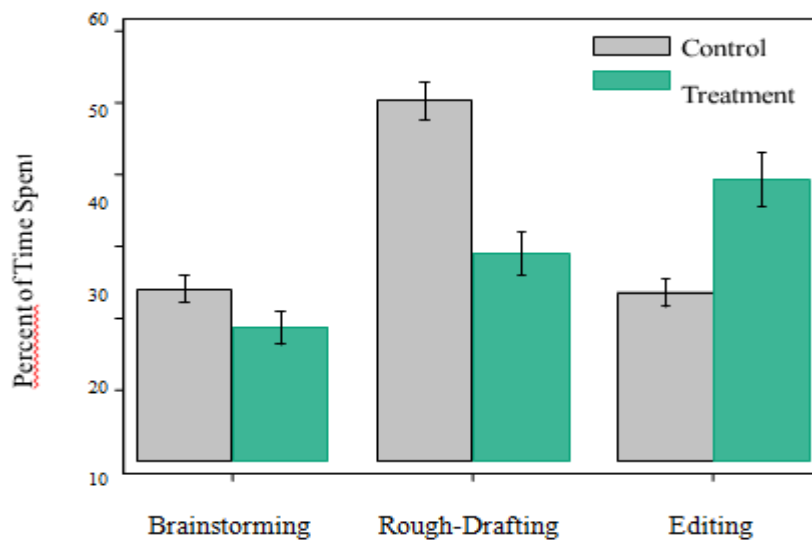
Skill Demand

If ChatGPT proves particularly beneficial to individuals with poor writing and communication skills relative to their other abilities, it could have significant labor market implications. This could expand occupational opportunities and increase earnings for those with strong idea-generation skills who struggle with effectively communicating those ideas. To test this hypothesis, we constructed two measures of a person’s relative writing skills. First, at the start of the experiment, we asked participants to rank their skills in communication (writing and speaking), problem-solving, and creativity from 1 to 3. Second, in addition to assigning overall grades, evaluators assessed each piece of output based on writing quality, content quality, and originality, allowing us to measure the gap between a participant’s first-task overall score and their writing score.



Similarly, we developed two measures of the individual-level benefits of ChatGPT. First, at the end of the experiment, we asked treatment-group participants how much they would be willing to pay monthly to access ChatGPT in their jobs. Second, we measured the increase in each treatment participant’s grade from the first to the second task. Our findings provide no clear evidence to support the hypothesis. Figure 3, Panel B, shows that the average willingness to pay for ChatGPT is consistent across the terciles of both writing skill measures: respondents, irrespective of their writing skills, are willing to pay about 0.5% of their monthly salary for a ChatGPT subscription. Additionally, the grade improvements from ChatGPT are relatively flat across both measures of writing skills, indicating that individuals with poorer writing skills do not experience disproportionately large grade gains.

(a) Effects on Task Structure



No Clear Heterogeneity in Benefits of ChatGPT by Relative Writing Skills

Outcome: Willingness to Pay for ChatGPT (% Salary)

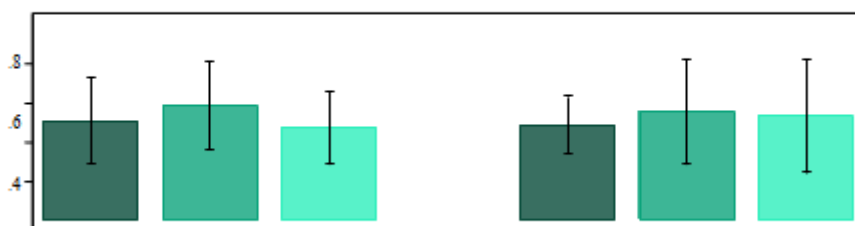


Figure 2: Effects on Task Structure and Skill Demand

V. DISCUSSION

The findings from our study reveal significant insights into the productivity impacts of ChatGPT on mid-level professional writing tasks, demonstrating both time efficiency and quality enhancements. The experimental design and supplementary interventions provide a comprehensive understanding of how generative AI can influence professional workflows. Our primary results indicate that ChatGPT substantially reduces task completion time by 37% and improves output quality, as evidenced by a 0.45 standard deviation increase in evaluator grades. These effects are robust across different incentive schemes, highlighting the versatility of ChatGPT in various motivational contexts. Furthermore, the distributional analysis shows that ChatGPT benefits workers across the board, particularly aiding those who initially performed poorly, thus reducing productivity inequality. The supplementary intervention requiring fixed time on tasks reinforces the notion that the quality improvements are directly attributable to ChatGPT’s



capabilities, independent of time saved. This suggests that ChatGPT enhances productive capacity by enabling higher quality output within a fixed effort framework. The post-task editing intervention underscores the dual utility of ChatGPT, where participants leveraged the tool to not only save time but also to refine and enhance the quality of their initial submissions. This dual utility underscores the broader potential of ChatGPT as a tool for both efficiency and creativity in professional settings. Our findings also highlight the increased job satisfaction and self-efficacy among users, pointing to the positive psychological impacts of integrating such AI tools in the workplace. Despite some control group participants using ChatGPT, thereby providing conservative estimates of its effects, the significant productivity gains observed suggest a profound potential for generative AI in transforming professional writing and other cognitive tasks. These results offer valuable implications for organizations considering the adoption of AI tools, suggesting that integrating ChatGPT can lead to substantial improvements in productivity, quality, and employee satisfaction. The study also raises important considerations about the future of work and the evolving role of automation technologies, urging policymakers and industry leaders to thoughtfully navigate the integration of AI to maximize benefits while mitigating potential concerns about automation.

## VI. CONCLUSION

Our study provides compelling evidence on the significant productivity benefits of generative AI, specifically through the use of ChatGPT, in mid-level professional writing tasks. The experimental design, which included a robust sample of 444 college-educated professionals across various occupations, allowed for a comprehensive analysis of ChatGPT's impact. The results demonstrate that ChatGPT not only reduces the time required to complete tasks by 37% but also significantly enhances the quality of the output, as reflected in a 0.45 standard deviation increase in evaluator grades. These improvements were consistent across different incentive schemes, reinforcing the tool's versatility and effectiveness. The supplementary interventions further elucidate the specific mechanisms through which ChatGPT operates, highlighting its ability to enhance productive capacity and output quality within a fixed effort framework. The post-task editing option revealed that users appreciate ChatGPT for both its time-saving capabilities and its potential to improve the quality of their work. This dual utility positions ChatGPT as a valuable asset in professional settings, capable of boosting efficiency and creativity. The study also underscores the positive psychological impacts of AI integration, with increased job satisfaction and self-efficacy among users. Despite some control group participants using ChatGPT, the conservative estimates still indicate significant productivity gains, suggesting that the true impact of ChatGPT could be even greater. These findings have profound implications for organizations considering the adoption of AI tools, highlighting the potential for substantial improvements in productivity, quality, and employee satisfaction. Moreover, the study prompts important considerations about the future of work and the role of automation technologies, urging policymakers and industry leaders to navigate the integration of AI thoughtfully to maximize its benefits while addressing potential concerns. As generative AI continues to evolve, its application in professional contexts promises to transform workflows, enhance productivity, and redefine the boundaries of human-AI collaboration.

## REFERENCES

1. Acemoglu, Daron and Pascual Restrepo, "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 2018, 108 (6), 1488–1542.
2. and , "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, 128 (6), 2188–2244.
3. Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb, "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction," *Journal of Economic Perspectives*, 2019, 33 (2), 31–50.
4. Autor, David, "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, 2015, 29 (3), 3–30.
5. and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 2013, 103 (5), 1553–1597.
6. Boustan, Leah Platt, Jiwon Choi, and David Clingingsmith, "Automation After the Assembly Line: Computerized Machine Tools, Employment and Productivity in the United States," NBER Working Paper, 2022.
7. Hoffman, Mitchell, Lisa B Kahn, and Danielle Li, "Discretion in Hiring," *Quarterly Journal of Economics*, 2018, 133 (2), 765–800.
8. Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe, "AI, Skill, and Productivity: The Case of Taxi Drivers," NBER Working Paper, 2022.
9. Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mul-lainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 2018, 133 (1), 237–293.





10. Lee, David S, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *The Review of Economic Studies*, 2009, 76 (3), 1071–1102.
11. Schwabe, Henrik and Fulvio Castellacci, “Automation, Workers’ Skills and Job Satisfaction,” *Plos one*, 2020, 15 (11), e0242929.
12. G. Krithiga, V. Mohan, G. Chitrakala & S. Senthilkumar, “Optimization of switching angles for selective harmonic elimination in cascaded h-bridge multilevel inverters employing artificial intelligence techniques – A mini review”, *International Journal of Engineering Technologies and Management Research*, vol. 10, no. 1, pp. 1–16, 2023. <https://doi.org/10.29121/ijetmr.v10.i1.2023.1278>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | [ijmrset@gmail.com](mailto:ijmrset@gmail.com) |

[www.ijmrset.com](http://www.ijmrset.com)