On the Economic Consequences of Artificial Intelligence Generative Tools

Francesco Losma

November 22, 2024

1 Introduction

Artificial Intelligence (AI) encompasses a range of technologies that enable computers to perform tasks traditionally limited to human capabilities. These tasks include recognising speech, text, and images; making recommendations; analysing vast amounts of data; and potentially much more, given the rapid advancement of these technologies. Popularised by large language models such as OpenAI's ChatGPT, Google's Gemini, or Microsoft's Copilot, Generative Pretrained Transformers (GPTs) represent a specific type of artificial intelligence called generative AI. Unlike other technologies that require substantial upfront investments, GPTs are relatively inexpensive, easily scalable at the company level, and capable of directly influencing everyday activities, such as drafting emails, organizing events, and brainstorming ideas (Bick, Blandin, and Deming, 2024). As a result, these tools have the potential to profoundly impact the labour market in ways that are different from earlier digitalisation waves (such as the adoption of computers, the internet, or industrial robots). However, while these earlier technologies predominantly affected mid-skill service workers (computers) or low-skilled workers (robots), artificial intelligence has the potential to impact a broader range of workers because of its versatility and adaptability. Unlike earlier technologies, which were often task-specific, generative AI can perform tasks that span cognitive, creative, and analytical domains, enabling their application in both routine and complex tasks across industries. However, researchers have not thoroughly assessed the extent of these effects.

2 Literature Review

To date, the economic literature offers limited direct evidence of AI's impact more broadly (or generative AI, more specifically) on labour dynamics. This section summarizes the key findings in this emerging area of research without claiming to be exhaustive or including all papers in this rapidly growing body of literature.

2.1 Understanding job exposure to AI

One strand of literature seeks to understand the extent to which different jobs are exposed to AI, similar to previous studies on earlier technological advancements (Frey and Osborne, 2017). Key examples include studies by Eloundou et al. (2023), Webb (2020), and Felten, Raj, and Seamans (2021). Eloundou et al. (2023) focus specifically on generative AI, evaluating occupations based on their alignment with the capabilities of large language models (LLMs). Their findings suggest that approximately 80% of the U.S. workforce could have at least 10% of their tasks affected by the introduction of LLMs, with higher-income and typically higher-skilled jobs facing greater exposure. Webb (2020) constructs a measure of task exposure to AI by analysing the overlap between job task descriptions and the text of AI-related patents. His findings suggest that AI has the potential to substitute tasks traditionally performed by highly skilled workers.

2.2 Models on the substitutability of workers by AI

Other studies, drawing on the standard skill-biased technical change literature (Autor and Dorn, 2013), have adopted a more theoretical approach to clarify the potential substitutability or complementarity between AI tools and workers with varying skills and characteristics. For instance, Bloom et al. (2024) employ a simple CES nested production function that incorporates three types of capital - standard capital, industrial robots, and AI - alongside high-skilled and low-skilled labour. By assuming that AI is more substitutable for high-skilled than low-skilled are for high-skilled workers, the CES production function immediately yields that AI adoption will reduce the skill premium, thereby lowering wage inequality. Given the importance of providing a concrete theoretical foundation for this issue, a separate section (not included here for the sake of brevity) will explore the models underlying this topic in greater detail.

2.3 Surveys of AI use at the worker level

Survey-based studies on this topic at the worker level are still in their early stages, with only a few surveys developed by researchers to explore questions such as who uses generative AI in practice, to what extent, and for which tasks. Notable and very recent examples of this literature include Bick, Blandin, and Deming (2024) and Humlum and Vestergaard (2024). In the first U.S.-based study, by incorporating questions on this topic into the Real-Time Population Survey, the authors find that in August 2024, approximately 39.4% of the U.S. working population aged 18-64 had used generative AI. Of those employed, 28% reported using generative AI at work, with the majority (24.2%) using it at least weekly, and 10.6% using it daily. Generative AI use is more prevalent outside of work, although less intensive. Additionally, the study shows that generative AI adoption is more common among younger, more educated, and higher-income workers. From a gender perspective, men are 9% more likely to use generative AI at work[1]. The use of generative AI is also more frequent in management, business, and computer-related occupations, aligning with findings from the studies referenced earlier. Using a Danish-based survey of approximately 100,000 workers across 11 occupations exposed to AI tools[2], Humlum and Vestergaard (2024) confirm these findings. About 50% of workers reported using ChatGPT, with higher usage among younger, less experienced, higher-achieving (more educated and with higher grades), and male workers (women are 20% less likely to use these tools, even within a restricted occupation-task cell). The authors also explore the barriers to more widespread adoption of these tools. The main obstacles reported are related to specific firm policies: 43% indicated they need training to use ChatGPT, while 35% mentioned that employers actively restrict its usage. "Existential fears" about job redundancy or over-reliance on technology were among the least significant barriers. The relevant frictions also differ by occupation. Employer restrictions are more likely to bind in occupations that handle sensitive information, like financial advisors or legal professionals. Less IT-prone occupations, such as teachers, report a greater need for training. Customer service representatives are particularly cautious about using ChatGPT due to fears of being replaced or becoming overly dependent. Lastly, in occupations where writing is a core competency, such as journalism and teaching, workers tend to resist ChatGPT because it diminishes their enjoyment.

[1] Carvajal, Catalina, and Isaksson (2024) confirm these findings through a survey of students at a leading business school in Norway. While they also observe that women are less likely than men to adopt generative AI, they further reveal that this gap is predominantly driven by women at the top of the skill distribution.

[2] These occupations include accountants, customer support specialists, financial advisors, HR professionals, IT support specialists, journalists, legal professionals, marketing professionals, office clerks, software developers, and teachers

2.4 Surveys of AI use at the firm level

At the firm level, the 2019 Annual Business Survey (ABS) includes a module investigating the use of five digital technologies, including AI, as reported by Acemoglu et al. (2024)[3]. This survey reveals that, over the three-year period from 2016 to 2018, only a small minority of firms (3.2%) adopted AI. However, since adoption is concentrated among larger firms, 12.6% of workers were employed at firms using AI. Naturally, adoption rates differ across industries, and within industries, larger and younger firms are more likely to adopt these tools[4]. Regarding the intended use, 55% of surveyed firms using AI cited automation as their primary motivation. Consistent with this rationale, firms adopting AI exhibit higher labour productivity and a lower labour share. Nevertheless, firms do not anticipate reducing employment levels, as AI adoption has also led to an increased demand for skills.

2.5 *Causal* impact of AI on worker-level outcomes

Some studies adopt a micro-perspective, analysing the impact of AI on individual worker-level productivity. Dell'Acqua et al. (2023) examine the productivity effects of large language models (specifically, ChatGPT powered by GPT-4) on a sample of 758 business consultants at Boston Consulting Group (BCG), who are typically high-skilled and high-wage individuals. They find that while not all tasks performed by business consultants can be easily automated using AI - and, in some cases, blind reliance on AI may lead to less accurate and imprecise business solutions - AI can have a positive impact on productivity for tasks that are well-suited to automation. Specifically, AI increases productivity both at the extensive margin (with an average of 12.2% more tasks completed) and the intensive margin (tasks are completed 25.1% faster on average). Further analysis of heterogeneous treatment effects shows that lower-performing consultants benefit the most, with average performance improvements of 43%, compared to only a 17% increase for those who are already highly productive. In a controlled lab experiment, Noy and Zhang (2023) examine the productivity effects of AI on relatively simple and clearly AI-feasible writing tasks. Their findings indicate that the productivity distribution compresses once AI use is permitted, as lower-ability workers benefit disproportionately, leading to a narrowing of productivity gaps. Similarly, Brynjolfsson Li and Raymond (2023) investigate the productivity impact of generative AI on a sample of over 5000 customer support agents - a role with strong overlaps with the capabilities of generative AI tools. They find a positive treatment effect, particularly for inexperienced and less-skilled workers. Notably, their study also reveals that AI usage improves employee

[3]This study builds upon the work using the 2018 Annual Business Survey (ABS), as summarised by Zolas et al. (2020).

[4]Focusing only on job vacancy data that require AI-related skills, and linking them with firm-level data from Compustat, Alekseeva et al. (2021) confirm that firms with larger market capitalizations, higher cash holdings, and greater investments in R&D exhibit a stronger demand for AI skills.

retention.

2.5.1 Role of AI in boosting innovation

Another key question is whether AI can enhance inventors' innovative efforts and be a new general-purpose technology (GPT) that reshapes the nature of the innovation process (Cockburn, Henderson, and Stern, 2018), for instance by changing the way in which hypothesis is generated (Ludwig and Mullainathan, 2024). Focusing on new materials discovery[5], Toner-Rodgers (2024) finds that generative AI increases new material discoveries by 44% and patent filings by 39%. Notably, unlike prior evidence on simpler tasks, while the bottom third gains minimal benefits, the output of top researchers nearly doubles. High-performing scientists excel at selecting the most promising AI-generated material candidates[6], whereas lower-tier scientists often get bogged down by testing false positives.

2.6 *Causal* impact of AI on firm-level outcomes

Other studies have used job vacancy data to infer firm-level adoption of artificial intelligence tools[7] and to examine the impact of these technologies on various firm outcomes. Notable examples of this literature include Copestake et al. (2024) - which use data on India - and Acemoglu et al. (2022) - which focus the US. In both studies, the authors use the frequency of hiring for AI-related roles as a proxy for AI adoption at the establishment level, investigating whether establishments with higher recruitment of AI-related workers simultaneously reduce hiring in non-AI positions. Despite the differing economic and institutional contexts of the United States *vis-à-vis* India, the findings are largely comparable. Copestake et al. (2024) find that firms adopting AI significantly reduce growth in non-AI job postings and offered wages, with this effect being particularly pronounced among high-skilled managerial and professional occupations, non-routine work, and analytical and communication tasks. These findings indicate marked differences compared to previous waves of digitalisation (Autor and Dorn, 2013). Similarly, Acemoglu et al. (2022) confirm these trends in the U.S. context, although they do not observe significant AI-labour substitution[8]. Using a similar approach, Babina

[5]Though highly insightful, I believe the study's external validity is somewhat limited. As the author notes, the process for discovering new materials is fairly standardised, involving domain knowledge and iterative computational procedures to generate initial designs. In this context, it is clear that a large language model, when properly prompted, can significantly accelerate this initial phase.

[6]Consistent with the literature that illustrates how new technologies give rise to new tasks (exemplified by Acemoglu, Kong and Restrepo (2024)), this is a clear example of a novel task introduced by AI: the evaluation of AI-generated material candidates.

[7]As note by Babina et al. (2024), a key challenge in studying the economic impact of AI technologies is the lack of comprehensive data on firm-level AI adoption.

[8]They also observe both a *displacement* and a *reinstatement* effect: once establishments adopt AI, certain skills that were previously in demand are no longer required, while new skills become sought after.

et al. (2024) study the impact of AI adoption on various firm-level outcomes, such as sales, employment, and market valuations, and find a positive impact across all these dimensions. Importantly, and related to the individual-level study of Toner-Rodgers (2024), their analysis reveals that AI drives firm growth primarily by enhancing product innovation, as evidenced by an increase in product patents and trademarks, while it does not appear to influence process innovation. Specifically, they do not detect any measurable impact on changes in sales per worker, total factor productivity (TFP), or labour displacement.

2.7 Studies of AI in developing settings

Most of the focus of this literature has been on developed countries, especially the US. Examples of studies investing the potential effect of various forms of artificial intelligence in developing countries are Copestake et al. (2024) - which was discussed above - and Otis et al. (2023) who construct a randomised field experiment designed to understand whether generative AI (GPT-4) can be used to assist entrepreneurs with their daily activities in Kenya. In general, they do not find any average treatment effect of using this AI assistant tool. However, when looking at the heterogeneous effect among different skilled entrepreneurs, they find that high performers entrepreneurs were able to increase their business performance by about 20%, whereas low performers decreased their business performance by 10%. Analysing the exact questions asked, they found that low performers were asking more challenging business questions. This resonates well with the findings of Dell'Acqua et al. (2023) that generative AI should not be considered a *panacea* capable of addressing all possible tasks.

3 Research Proposal and Empirical Design

Building on the existing literature, I believe there are several promising avenues for further research in this area. One project I propose involves integrating these diverse strands by rigorously identifying the causal effects of adopting generative AI tools on both firm-level and worker-level outcomes together. Analysing the labour market outcomes for workers differentially affected by the introduction of AI is essential. For example, the literature currently lacks a thorough investigation of how AI adoption influences the career trajectories of impacted workers. Another crucial, yet underexplored, aspect concerns the *displacement effect* of AI. Specifically, if displacement occurs, which groups are most affected? Toner-Rodgers (2024) tentatively addresses this question, finding that, following his experiment, the laboratory where it was conducted fired 3% of its researchers - 83% of whom were in the bottom quartile for their ability to assess innovations generated by AI. Directly observing this link is crucial, as it could exacerbate existing inequalities (e.g., by gender).

Building on previous research I have done on telework, another key question I aim to address is whether high-skilled jobs in developing countries could face competition from mid-skilled workers in other developing countries, once the latter are equipped with AI and teleworking tools. Analysing export patterns from service companies in a developing country to a developed country after the introduction of AI could provide a valuable way to test this hypothesis.

My proposed research involves implementing a randomized controlled trial (RCT) within a developing context, supplemented by survey waves to capture key contextual factors, such as economic conditions, worker characteristics, and perceptions of AI adoption, which will enhance the understanding of the experiment's effects and the environment in which it is run. Previous studies examining earlier waves of digitalisation faced significant barriers due to high implementation costs. However, generative AI is uniquely accessible to firms at a relatively low cost, providing an opportunity to study its adoption and effects in a scalable manner. Drawing a parallel between AI adoption and the adoption of management practices, several studies have used randomized experiments to evaluate the effectiveness of such interventions (e.g., Bloom et al., 2015). Another relevant article in this literature is the management study by Bloom and colleagues (2013), which examines whether differences in management practices drive productivity disparities. They conducted a field experiment on large Indian textile firms, providing free consulting on management practices to randomly selected treatment plants. The results showed that treated firms improved product quality, enhanced assembly line efficiency, and reduced inventory stocks, leading to a 17% increase in productivity. This paper is particularly relevant for the econometric specifications employed, as my proposed research is also likely to involve a relatively small sample of treatment and control establishments[9]. Their experiment included only 28 plants across 17 firms, raising several important concerns: (i) whether the sample size is sufficient to identify significant impacts, (ii) which type of statistical inference is appropriate in this context, and (iii) whether the small sample can adequately represent large firms in developing countries more broadly. While concerns about representativeness are common to all types of experiments, their approach to addressing the issues of sample size and statistical inference involved collecting very high-frequency observations. Specifically, they recorded weekly data on their outcomes of interest, allowing them to use estimators that rely on a large number of time periods (T) rather than a large number of cross-sectional units (N). This approach mitigates the limitations of a small sample size, and I plan to implement a similar strategy in my study, collecting frequent observations to enhance the precision of my estimates.

[9]While this limitation will not pose significant challenges for the worker-level analysis, it is more problematic for addressing establishment-level research questions.

The choice of a developing context - specifically, India - is particularly important, as such countries continue to experience a productivity disadvantage relative to more developed nations. Understanding whether these new waves of digitalisation, which are relatively cheaper and more scalable than previous technologies, can help bridge this productivity gap is therefore of crucial policy importance. India serves as an ideal case study, given its large, service-based economy and its status as a major exporter of services (as highlighted in this Goldman Sachs Report). Moreover, the country features a relatively advanced IT infrastructure[10] and a strong supply of technical IT skills (the share of job vacancies specifying AI skills is higher in India than in Western countries like the US, UK, or Canada - as shown by Copestake et al., 2024) - all factors that are fundamental to sustain the use of artificial intelligence tools.

The aim is to contact a large, multi-establishment firm that stands to benefit significantly from introducing generative AI into their daily operations. One potential candidate for this study could be a law firm operating across multiple cities, with a comparable hierarchical structure and consistent skills and tasks performed within each establishment. Law firms are particularly well-suited for this experiment due to their reliance on knowledge-intensive work, such as document review, legal research, and case analysis—tasks that can potentially be streamlined or enhanced by generative AI tools. Moreover, because of privacy reasons, law firms often block the adoption of generative AI tools (Humlum and Vestergaard, 2024). Therefore, this limits the contamination of the treatment in the control group and the non-compliance of the control group. Additionally, due to these sensitivity issues, the company may be more keen to partner up in a controlled experiment. These establishments will be randomised into three groups: a control group where generative AI will not be implemented, a treatment group where generative AI will be introduced, and a second treatment group where generative AI will be introduced alongside training programs to equip workers with the skills necessary to effectively use these technologies. Indeed, some literature highlights the importance of teaching middle-skill workers how to use AI tools productively and efficiently (Autor, 2024).

All in all, looking across establishments could help clarify more macro-level issues. For instance, do treated firms export more services, which could help address whether AI and telework together might substitute high-skilled workers in developed countries with relatively lower-skilled workers in developing countries? What are the overall labour effects of this technology? On the other hand, comparing workers across establishments[11] could provide a more precise understanding of AI's productivity effects. From a broader perspective, this setting could also help clarify which types of workers are likely to benefit from or lose out to this technology, with particular attention to gender as a crucial factor.

[10] Source: CRN India

[11] The underlying assumption here is that establishments within the same firm are comparable, and workers across different establishments who perform the same tasks but are differentiated only by location are also comparable.

Once collaboration is established, several more questions could be answered. For instance, another interesting question is whether AI tools - mostly trained on data generated in the developed world - can be useful in a developing context *off-the-shelf*, or whether a tailored large language model (LLM) based on the specific needs of a company in a developing world would greatly outperform standard generative AI tools. This is particularly relevant, as not all companies have the resources to train a model *in-house*, which could exacerbate inequalities across companies and further increase productivity gaps. Another compelling question is whether developing countries are equipped with the skills to fully exploit this technological revolution. As Acemoglu and Zilibotti (2001) clearly demonstrate, technologies developed in the *north* may be less productive when used in the *south* due to a lack of the necessary skill set.

Bibliography

Acemoglu, D., Kong, F., and Restrepo, P. (2024). *Tasks at Work: Comparative Advantage, Technology and Labor Demand*. Working Paper 32872. National Bureau of Economic Research.

Acemoglu, D. and Zilibotti, F. (2001). "Productivity Differences". In: *The Quarterly Journal of Economics* 116.2, 563–606.

Acemoglu, D. et al. (2022). "Artificial Intelligence and Jobs: Evidence from Online Vacancies". In: *Journal of Labor Economics* 40.1, pp. 293–340.

Acemoglu, D. et al. (2024). *Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey*. Working Paper 30659. National Bureau of Economic Research.

Alekseeva, L. et al. (2021). "The Demand for AI Skills in the Labor Market". In: *Labour Economics* 71.

Autor, D. (2024). *Appplying AI to Rebuild Middle Class Jobs*. Working Paper 32140. National Bureau of Economic Research.

Autor, D. H. and Dorn, D. (2013). "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market". In: *American Economic Review* 103.5, 1553–97.

Babina, T. et al. (2024). "Artificial Intelligence, Firm Growth, adn Product Innovation". In: *Journal of Financial Economics* 151, p. 103745.

Bick, A., Blandin, A., and Deming, D. J. (2024). *The Rapid Adoption of Generative AI*. Working Paper 32966. National Bureau of Economic Research.

Bloom, D. E. et al. (2024). *Artificial Intelligence and the Skill Premium*. Working Paper 32430. National Bureau of Economic Research.

Bloom, N. et al. (2013). "Does Management Matter? Evidence from India". In: *The Quarterly Journal of Economics* 128.1, 1–51.

Bloom, N. et al. (2015). "Does Working from Home Work? Evidence from a Chinese Experiment". In: *The Quarterly Journal of Economics* 130.1, pp. 165–218.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). *Generative AI at Work*. Working Paper 31161. National Bureau of Economic Research.

Carvajal, D., Catalina, F., and Isaksson, S. (2024). *Will Artificial Intelligence Get in the Way of Achieving Gender Equality?* Discussion Paper 03. NHH.

Cockburn, I. M., Henderson, R., and Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation*. Working Paper 24449. National Bureau of Economic Research.

Copestake, A. et al. (2024). *AI and Service-Led Growth: Evidence from Indian Job Adverts*. Working Paper. University of Oxford.

Dell'Acqua, F. et al. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*. Working Paper 24-013. Harvard Business School.

Eloundou, T. et al. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. arXiv: 2303.10130.

Felten, E., Raj, M., and Seamans, R. (2021). "Occupational, Industry, and Geographic £xposure to Artificial Intelligence: A Novel Dataset and its Potential Uses". In: *Strategic Management Journal* 42.12, pp. 2195–2217.

Frey, C. B. and Osborne, M. A. (2017). "The Future of Employment: How Susceptible are Jobs to Computerisation?" In: *Technological Forecasting and Social Change,* 114, pp. 254–280.

Humlum, A. and Vestergaard, E. (2024). *The Adoption of ChatGPT*. Discussion Paper 16992. IZA.

Ludwig, J. and Mullainathan, S. (2024). "Machine Learning as a Tool for Hypothesis Generation". In: *The Quarterly Journal of Economics* 139.2, pp. 751–827.

Noy, S. and Zhang, W. (2023). *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*. Working Paper. MIT.

Otis, N. et al. (2023). *The Uneven Impact of Generative AI on Entrepreneurial Performance*. Working Paper 24-042. Harvard Business School.

Toner-Rodgers, A. (2024). *Artificial Intelligence, Scientific Discovery, and Product Innovation.*

Webb, M. (2020). *The Impact of Artificial Intelligence on the Labor Market*. Job Market Paper. Stanford University.

Zolas, N. et al. (2020). *Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey*. Working Paper 28290. National Bureau of Economic Research.