Department of Economics, University of Warwick

Monash Business School, Monash University

as part of

Monash Warwick Alliance

The Effect of Generative AI Adoption on Knowledge Workers: Evidence from Luxembourg

Sofiya Musina

Warwick-Monash Economics Student Papers

September 2024 No: 2024/75

ISSN 2754-3129 (Online)

https://warwick.ac.uk/fac/soc/economics/research/wmesp/

The Warwick Monash Economics Student Papers (WM-ESP) gather the best Undergraduate and Masters dissertations by Economics students from the University of Warwick and Monash University. This bi-annual paper series showcases research undertaken by our students on a varied range of topics. Papers range in length from 5,000 to 8,000 words depending on whether the student is an undergraduate or postgraduate, and the university they attend. The papers included in the series are carefully selected based on their quality and originality. WM-ESP aims to disseminate research in Economics as well as acknowledge the students for their exemplary work, contributing to the research environment in both departments.

Recommended citation: Musina, S. (2024) The effect of generative AI adoption on knowledge workers: evidence from Luxembourg. *Warwick Monash Economics Student Papers* 2024/75.

[1] Warwick Economics would like to thank Gianna Boero and Samuel Obeng for their contributions towards the selection process.

The Effect of Generative AI Adoption on Knowledge Workers: Evidence from Luxembourg

Sofiya Musina*

Abstract

Large Language Models (LLMs), such as ChatGPT, demonstrate an unprecedented applicability in a variety of domains, and, unlike previous waves of innovation, are capable of nonroutine cognitive tasks - leaving educated, white-collar workers most exposed. However, few studies address the relevant labour market implications outside controlled experimental environments. This project investigates the effects of LLMs on knowledge worker competency requirements using a difference-in-difference model based on a sample of 105,912 online job advertisements (Luxembourg, 2020-2024). The findings contain weak evidence that LLMs cause a reduction in demand for experience, education, cognitive skills and creativity, while leaving soft skills unaffected.

Keywords: Employment, Skills Demand, Technology, AI

JEL Classifications: J01, J23, J24, O33

---

* **Contact email:** sofiya.musina@gmail.com

Appendix available online at: https://github.com/sofiyamusina/ai_and_jobs_ec331/blob/main/WM-ESP%20Dissertation%20-%20Appendix.pdf *(note: click "Download Raw file" to access links in the Appendix)*

Acknowledgements: I would like to thank Natalia Zinovyeva, my supervisor, for her guidance along the way as she helped me channel my interest in AI into an economic research project. In addition to sharing valuable thoughts, she made the process of writing itself enjoyable and full of learnings. I would also like to thank my friend, Reem Jaiswal, for her support during this journey.

1 Introduction

ChatGPT is an Artificial Intelligence (AI) text generator based on an LLM (Large Language Model). Released in November 2022, ChatGPT reached 100m weekly users in less than a year (Porter, 2023) - demonstrating just how quickly Generative AI is being adopted.

Unlike other AI systems that are domain-specific (Marr, 2023), LLMs are capable of multi-subject knowledge and information processing, thus can be applied in a variety of contexts (Eloudou et al., 2023). Such vast general purpose applicability could mean potentially extensive, profound changes in labour markets, which calls for re-evaluating some of the old concerns about AI and workers. Does AI replace workers, increase their productivity or create new jobs? Which factors determine whether workers are shielded from, at risk of, or benefit from AI? Currently, little research exists addressing these issues from an LLM perspective, with no clear consensus - this is why the effects of LLMs on labour markets are the focus of this project.

This fits into the wider academic context of research on skill-biased technological change: the idea that technological advancements cause shifts in labour demands (Autor et al., 2003). The most recent substantial technology-driven shift in labour markets occurred in 1960-2000 as a result of computerization, which reduced the demand for routine (automatable) labour and increased demands for nonroutine (non-automatable) labour, thus increasing demand for educated, “knowledge” workers who held a competitive advantage in the latter (Autor et al., 2003). Today, the ability of AI, specifically LLMs, to carry out many “nonroutine” tasks (Felten et al., 2021; Felten et al., 2023) suggests different consequences of skill-biased technological change, especially for “knowledge” workers.

Unlike the research covering labour market implications of AI in general, the majority of LLM-focused research consists of RCTs with “ChatGPT” and “no ChatGPT” conditions. Examples include Dell’Acqua et al. (2023) assessing the extent to which ChatGPT increased productivity of consultants, or Noy and Zhang (2023) investigating whether ChatGPT use drives productivity improvements in writing tasks. While high in internal validity, these studies are not easily generalizable to real-life labour markets. The approach in Hui et al. (2023) is higher in external validity: they examine employment outcomes in online freelance platforms after ChatGPT release using difference-in-differences, with the drawback being that the "treatment" group consisted of writing-intensive jobs (copywriting, editing) only.

This project aims to match the external validity of Hui et al. (2023), but generate better generalizability by considering a wider array of occupations among the treated and controls. Additionally, instead of focusing on macroeconomic outcomes such as wages and employment, this project adopts a novel angle by focusing on the effects that LLMs might have specifically on the labour demand side, and specifically on worker skillsets and competencies.

This paper investigates the effects of LLM adoption on "knowledge" workers - those who benefited most from the 1960-2000 wave of computerization. More precisely, the research question is: **How are recent changes in demand for knowledge worker competencies influenced by occupational exposure to LLMs?**

A difference-in-differences model is estimated on a dataset of 105,912 job ads scraped from 8 websites in Luxembourg by Techmap.io over 2020-2024. The variables of interest - knowledge worker competency demands - are extracted from the text of job ads using language processing. The treated group is assigned based on the advertised job's LLM exposure, and the post-treatment period constitutes the period after ChatGPT release in November 2022. A difference-in-difference method is used to mitigate selection bias regarding initially different competency demands in occupations with different LLM exposures (e.g. lawyers and programmers). The main assumption necessary for causal interpretation is parallel trends, which holds if the demands for any specific worker competency evolve similarly over time for different occupations with different LLM exposures in the absence of the treatment ("ChatGPT shock").

Hypothetically, LLMs can affect labour competency demands through various mechanisms. First, in a purely technical, production function manner, AI functionality can either substitute a competency, reducing the demand for it, or complement/augment the competency, boosting demand. Another impact channel is labour supply elasticity: AI can make worker upskilling easier and cheaper, increasing skill supply, contributing to lower wages (lower costs for employers), and therefore increased demand for the skill. Additionally, such AI-driven reductions in skill scarcity can boost output of firms or industries demanding the skills, increasing relevant worker demand in the growing firm or industry, thus increasing employment and wages. All these processes have different, sometimes ambiguous, implications for wages and employment, and likely occur on different time horizons.

Such ambiguity makes it challenging to interpret estimation results and extrapolate them. Methodological issues (attenuation, noise, etc.) only added to the challenge. Overall, the estimated model has not identified LLM effects on demands for 6 competencies, and demonstrated weak evidence for negative effects on demands for experience, education, cognitive skills and creativity. A production function substitution effect is chosen as the most plausible explanation, implying possible negative wage and employment effects for educated, experienced, creative and more intellectually able workers.

Moreover, one should be cautious about interpreting the findings of this study, as its low power remains a significant shortcoming. This is due to a relatively small dataset from a small labour market (Luxembourg), and the presence of multiple testing: estimating multiple specifications with multiple parameters and dependent variables increases the likelihood of finding at least one statistically significant effect purely by chance. Overall, this presents the risk of overstating LLM impacts.

2 Literature Review

The most influential research in the area of skill-biased technological change is Autor et al. (2003). Having observed that computer adoption is associated with increased demand for college-educated labour, the authors argue that the causal mechanism behind this process is the fact that computer capital substitutes for routine tasks while complementing cognitive non-routine tasks. Creating a theoretical model that incorporates this into a production function, they used 1960-1998 US Census data to validate this framework. A regression of changes in industry task input on industry computerization showed that computerization favoured educated workers, as they held an advantage in nonroutine tasks.

Would the effects on AI adoption today be the same, given that AI excels at non-routine cognitive tasks (OpenAI, 2023)? To investigate this, Felten et al. (2021, 2023) develop AI exposure scores for all US occupations, matching detailed task breakdowns of 774 occupations with tasks that AI can and cannot do. Similar to measures developed by Webb (2020) and Brynjolfsson et al. (2018), these scores measure the extent to which a job can be carried out by AI - crucially, without implying neither substitution nor complementarity. Although purely theoretical, Felten's scores provide a crucial measure of job exposure to academics interested in AI effects. These exposure scores reveal that jobs that are compensated better, require college degrees, involve information processing, writing or programming - in other words, "white collar" jobs - are most exposed to AI and LLMs (Felten et al., 2021, 2023), which means that educated labour might not benefit from mass AI adoption in the same way it did from computerization.

Acemoglu et al. (2022) use 2010-2018 US job posting data to create one of the most thorough existing studies of AI adoption effects in real labour markets. A regression of change in employment outcomes on Felten, Webb and Brynjolfsson scores reveals the association of AI exposure with a declining demand for some skills (e.g. customer services, maintenance), emergence of new skill demands (e.g. analysis, marketing, finance, IT), and less overall hiring. This highlights that the factors determining the direction of AI effects on labour demands are much more complex than whether the labour is "routine".

Unfortunately, Acemoglu et al. (2022) do not account for the recent advancements in Generative AI and LLMs. Among the few studies from the post-ChatGPT period, a notable one is Dell'Acqua et al. (2023) RCT involving BCG consultants. Participants, assigned to a "GPT access" or "no access" condition, carried out realistic, complex and knowledge-intensive tasks, applying ChatGPT outside its direct competency of text generation. LLM use improved worker productivity on creative and ideation tasks (+25.1% speed, +40% quality). Additionally, in these tasks low performers experienced 26% bigger gains from AI use - and one could extrapolate to suggest that less educated workers could be the ones benefiting from Generative AI, as it could help them reduce the gap to more educated workers. This is in direct contradiction with Autor et al. (2003) findings. Finally, analytical and data assessment task output was only enhanced by LLM use when participants acknowledged AI's limits and applied critical judgement instead of "blind" reliance. This suggests that critical thinking and some complex, interactive cognitive tasks cannot yet be substituted by LLMs.

Hui et al. (2023) use difference-in-differences to model the effect of ChatGPT release on outcomes of online freelancers, and observe a 2% fall in employment and 5.2% fall in earnings for freelancers in highly-exposed occupations, suggesting a substitution effect. However, the treatment group (high AI exposure) consisted of writing-intensive jobs (e.g. copywriting, editing) only, therefore these findings cannot be generalised to all "knowledge" workers.

Crucially, most studies on AI and labour do not find any conclusive impacts of AI on economy-wide wages, employment or productivity (Acemoglu et al., 2022; Babina et al., 2022). This might reflect the recency of AI technologies, and lags in their adoption or impact (Babina et al., 2022). Today, this recency issue is a major complication in research about AI and jobs.

Moreover, there remain numerous gaps. Autor et al. (2022), presenting conclusive evidence on AI effects, does not account for the recent advancements in and diffusion of Generative AI. Dell'Acqua et al.'s (2023) RCT, whilst high in internal validity and investigating ChatGPT directly, does not reveal whether any of the observed processes are reflected in firms' skill demands. Hui et al. (2023) cannot be generalised to all "knowledge" workers.

This project contributes to existing literature by attempting to address the above gaps, by focusing on the impacts of LLMs in "real-life" jobs markets, and by using a sample with a wide array of knowledge workers for generalizability. Additionally, a novel angle is defining some of the outcome variables as worker "competencies", or qualities, such as analytical, communication, leadership skills, etc. - rather than domain-specific knowledge (e.g. finance skills, marketing skills in Autor et al., 2022).

The following hypotheses serve as a starting point:

1. **Experience.** LLMs reduce demand for experienced labour: experience would lose value if AI helps access the knowledge that is otherwise accumulated through experience. The opposite effect of increased demand is possible if some worker qualities gained through experience cannot be substituted by AI, and experience gains relative value.
2. **Education.** LLMs reduce demand for university-educated labour, as they reduce the need for workers to have subject-specific knowledge and increase the accessibility of the information usually obtained at university.
3. **Cognitive skills.** LLMs increase demand for cognitive skills such as analysis, problem-solving and decision-making, because AI appears complementary to critical thinking (Dell'Acqua et al., 2023).
4. **Creativity.** LLMs reduce demand for creativity, as content generation and ideation are their strongest capabilities (Dell'Acqua et al., 2023).
5. **Soft skills: interpersonal, leadership, time management, independence, flexibility.** Since soft skills derive from human traits that are not AI-substitutable, LLMs either have no effect on soft skill demands, or have a positive effect on demand if soft skills gain value relative to other AI-substitutable competencies.

3 Data and Methodology

3.1 Data

3.1.1 Outcome and Treatment Variables

The main dataset used is a cross-section of 105,912 online job ads from Luxembourg in January 2020 - January 2024, scraped by Techmap.io, a hiring platform startup (see Appendix A). For each vacancy, information is available on the name, date posted, website posted on, contract type, text of posting, etc.

Some variables required additional work in order to represent appropriate measures of the dependent and independent variables. For instance, all the outcome variables were constructed from ad texts using keyword search (Appendix D). As a result, 19 variables were defined using 10 worker competencies: **experience, education, cognitive, creativity, interpersonal, leadership and time-management skills, motivation, flexibility and independence.** Except for education, all competencies were measured along the **extensive** and **intensive** margin, as defined in the below example:

Table 1. Outcome variable definition

Variable	Unit of measurement	Definition
Competency X, extensive margin	Probability	Dummy: 1 if Competency X required in job ad ***i***, 0 if not
Competency X, intensive margin	Number of keyword occurrences	Number of times Competency X related keywords appear in job ad ***i***

The treatment variable - LLM exposure - was not contained in the Techmap.io data, as the only currently available database detailing exposure of occupations to LLMs was created in Felten et al. (2023). They calculate LLM exposure scores for all US occupations by matching detailed task breakdowns of 774 standardised occupations (SOs) with tasks that LLMs can and cannot do. These scores measure the extent to which a job can be carried out by an LLM.

Table 2. Excerpt from Felten et al. (2023):
Full list of occupations sorted by language modelling exposure

Rank	SO code	Occupation title	Language modelling exposure
1	41-9041	Telemarketers	1.926
26	13-1071	Human Resources Specialists	1.557
58	41-3021	Insurance Sales Agents	1.427
154	17-2011	Aerospace Engineers	1.117
220	19-1021	Biochemists and Biophysicists	0.794
328	29-1141	Registered Nurses	0.272
376	39-5094	Skincare Specialists	-0.037
538	53-1011	Aircraft Cargo Handling Supervisors	-0.751
677	49-9041	Industrial Machinery Mechanics	-1.217

The 774 standardised SOs in Felten's list do not match perfectly to vacancy names in the Techmap data. To assign LLM exposure to vacancies, a matching algorithm was performed, inspired by the approach described by Otubusen and Sleeman (2021). Each vacancy name from the Techmap dataset was matched to a SO name from Felten's list, which allowed the assignment of an LLM exposure score to each vacancy based on the matched SO. The matching procedure consisted of vectorizing vacancy names and SOs using DistilBert, a pre-trained AI language model, and then creating a vacancy-SO match based on maximising cosine similarity (see Appendix B).

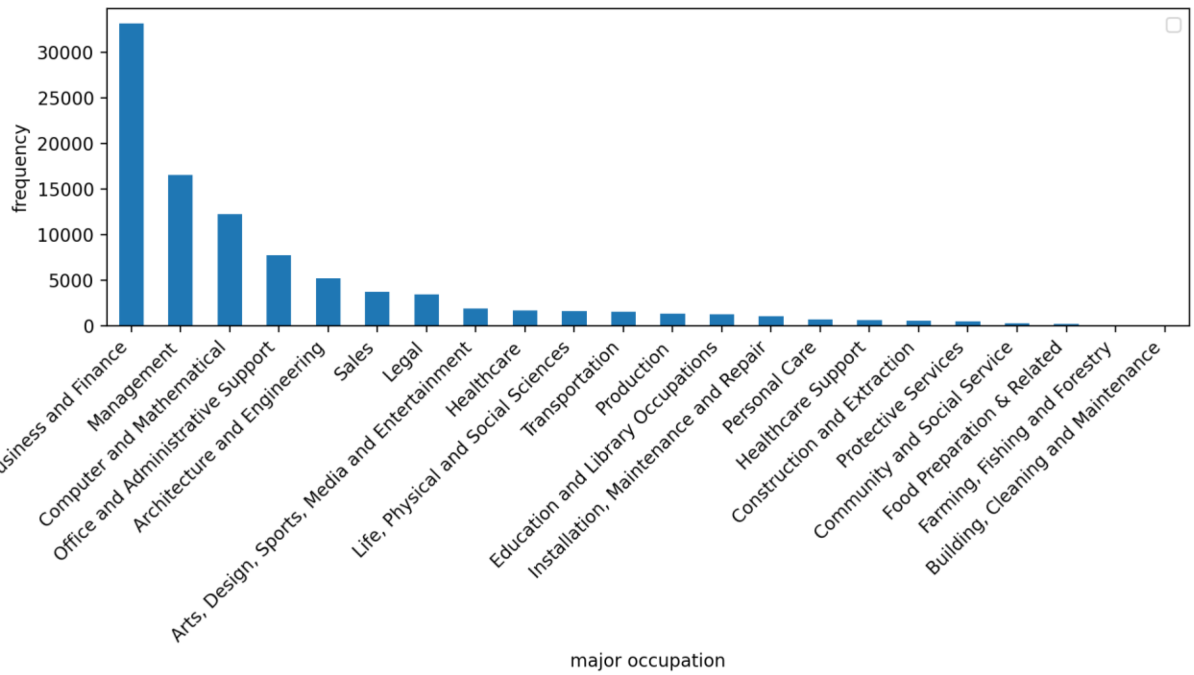
Table 3. Excerpt from the matching results

From Techmap.io data	From Felten et al. (2023)	
Vacancy name	**SO name**	**LLM exposure**
Sales & Administrative Assistant (m/f)	Sales managers	1.294
IT Analyst Developer - Application Integration	Software developers, applications	0.882
System & DevOps engineer (m/f)	Software developers, system software	1.166
Tax compliance officer: comfortable position	Compliance officers	0.572
Senior funds lawyer: transforming law firm	Lawyers	1.454
Account Payable Specialist	Financial specialists, all other	1.253
Accounting Manager - RE Investment firm	Financial managers	1.295
Luxembourg - Service Architect (F/M)	Architectural and engineering managers	0.647

3.1.2 Descriptive Statistics

The sample consists of vacancies for "white collar" or knowledge workers - those who, according to Felten et al. (2021, 2023) are most exposed to different forms of AI. Finance, Business, Management and Computer occupations are most widely represented in the sample:

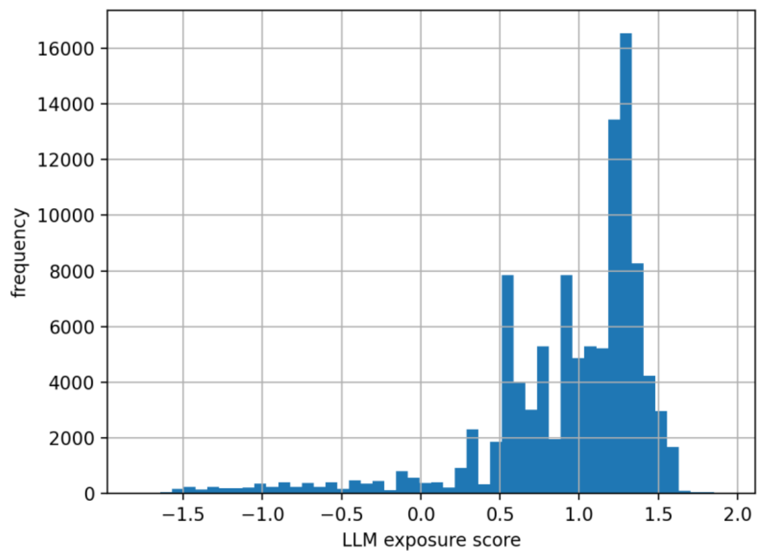
Image 1. Major occupation groups in sample



This is reflective of Luxembourg's labour market. The country's economy is heavily service-reliant, with 53% of the workers employed within finance, insurance, the public sector, education, human health, information, communications or technology (Koulischer et al., 2022).

Therefore, as expected, the distribution of LLM exposure in the sample is skewed to the right:

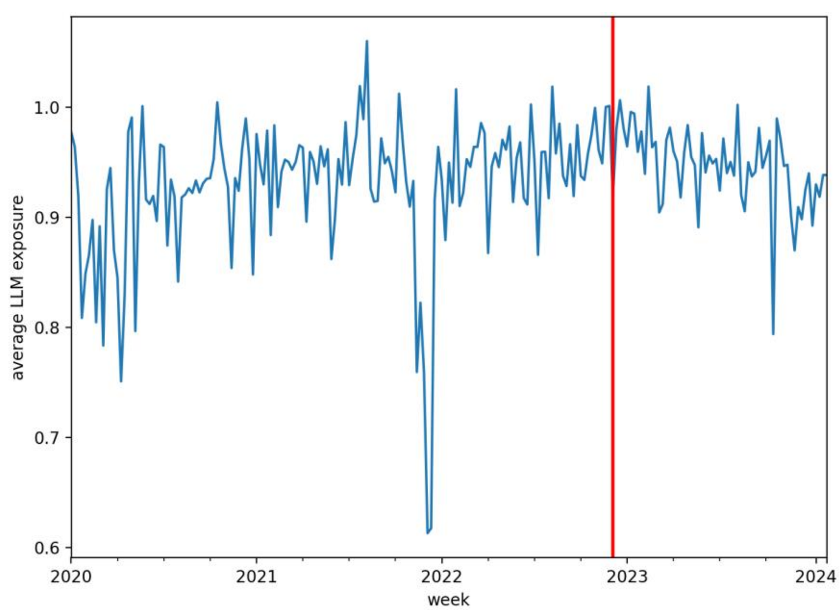
Image 2. LLM exposure distribution in the sample



This means that any analysis of this sample is only generalisable to high-AI-exposure, white-collar workers - who, luckily, represent the primary group of interest regarding LLM effects.

A plot of the advertised jobs' LLM exposure over time reveals a small-scale reversal from a positive trend to a negative one after the "ChatGPT shock":

Image 3. Weekly mean of LLM exposure in the sample



This could be interpreted as evidence for reduced demand for LLM-exposed vacancies, suggesting an overall substitution effect of LLMs on white-collar work. However, there are numerous other plausible explanations, including changes in source composition over time, adjustment of job markets after COVID, varying economic conditions altering hiring needs in different industries, or random shocks. Such ambiguity motivates the use of more advanced methods, such as difference-in-differences.

3.2 Methodology

The effect of interest is estimated using dynamic difference-in-differences, which allows to separately identify selection (the differences between AI-exposed and non-exposed vacancies' competency requirements due to different nature of the jobs) and the difference due to the treatment ("ChatGPT shock").

2 model specifications are estimated taking the below form:

$$\boxed{Y_i = \psi_{occ} + \delta_{q,q\neq 11} + \sum_{q=1,q\neq 11}^{17} \left(\beta_q \times LLM\_exposure_i \times D_q\right) + \gamma \times X'_i + \varepsilon_i}$$

In the above, i represents the unit of observation - a vacancy posting. Y_i is one of the outcome variables measuring worker competencies specified in Table 1; separate equations are fitted for each outcome variable. ψ_{occ} represents occupation fixed effects that control for the variation in education, experience and skill requirements across occupations. Vacancies matched to the same SO share the same ψ. δ_q is a dummy for each of 17 quarters except the baseline, controlling for time fixed effects (e.g. economic cycles). X'_i is a vector of controls, featuring the vacancy source. ε_i is the error term containing unobservables such as industry effects, firm effects, and random noise.

$\sum_{q=1,q\neq 11}^{17} \left(\beta_q \times LLM\_exposure_i \times D_q\right)$ is an interaction of treatment (LLM exposure) with quarterly dummies, estimating the dynamic treatment effect which is the primary effect of interest in this project. q=11 is the omitted baseline period (pre-ChatGPT shock quarter).

The sole difference between the specifications is the way $LLM\_exposure_i$, the treatment variable, enters the equation, and therefore how β_q, the coefficient of interest, is interpreted.

In **specification (1)**, $LLM\_exposure_i$ is a binary dummy defining a treatment (exposed) and a control (non-exposed) group based on a cutoff AI exposure score (see Appendix C):

$$\boldsymbol{LLM\_exposure_i} = \begin{cases} \mathbf{1}\ if\ AI\ exposure\ score\ i \geq cutoff \\ \mathbf{0}\ otherwise \end{cases}$$

Here, β_q estimates the **ATE (average treatment effect of the treated)**, and is interpreted as the difference in "ChatGPT shock" effects in quarter q (relative to q=11) between exposed and non-exposed occupations.

In **specification (2)**, $LLM\_exposure_i$ is a continuous treatment variable that measures treatment intensity. The more exposed an occupation is to LLMs, the more intensely it responds to the "ChatGPT shock". Here, β_q estimates the **ACR (average causal response)** - the average change in "ChatGPT shock" effects in quarter q (relative to q=11) as LLM exposure is increased by 1 unit.

In both (1) and (2), β_q identifies the effect of interest: the relationship between LLM exposure and worker competency demands. Correct identification relies on the parallel trends assumption, which, conceptually, holds if the demands for any specific worker competency would evolve similarly over time for different occupations in the absence of the "ChatGPT shock". (1) and (2) imply slightly different parallel trends. In (1), the "exposed" and "non-exposed" occupations need to follow the same trend in worker competency demands. (2), however, requires "strong parallel trends": the evolution of worker competency requirements for occupation *x* should be equal to the average evolution of worker competency requirements in all other occupations in the hypothetical case they experienced the "ChatGPT" shock with the same "intensity" as occupation *x*. Essentially, this is a restriction on treatment effect heterogeneity (Callaway et al., 2024).

The distinction between specifications (1) and (2) is arbitrary. In (1), the mean impact on more or less affected occupations is compared to arrive to the ATE. Specification (2) is the "restricted" version, as it imposes an additional assumption of a linear increase in the impact depending on LLM exposure – to arrive at the ACR. Both functional forms are explored because no previous research has indicated which of the two is more appropriate for this relationship.

Additionally, a dummy variable representing the website from which the job ad was extracted ("source") is included as a control. It reflects the time variation in the proportion of vacancies scraped from different websites, given the difference in vacancy AI exposure between the sources:

Image 4. Monthly number of vacancy ads per website in sample

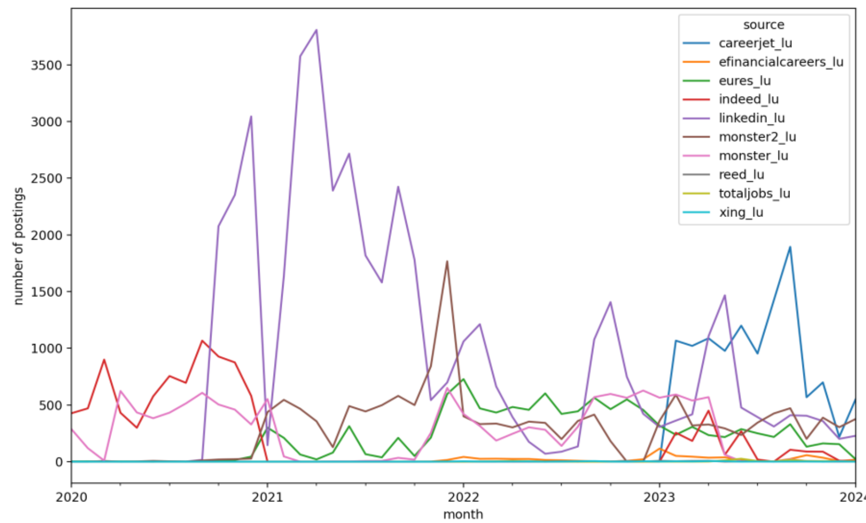
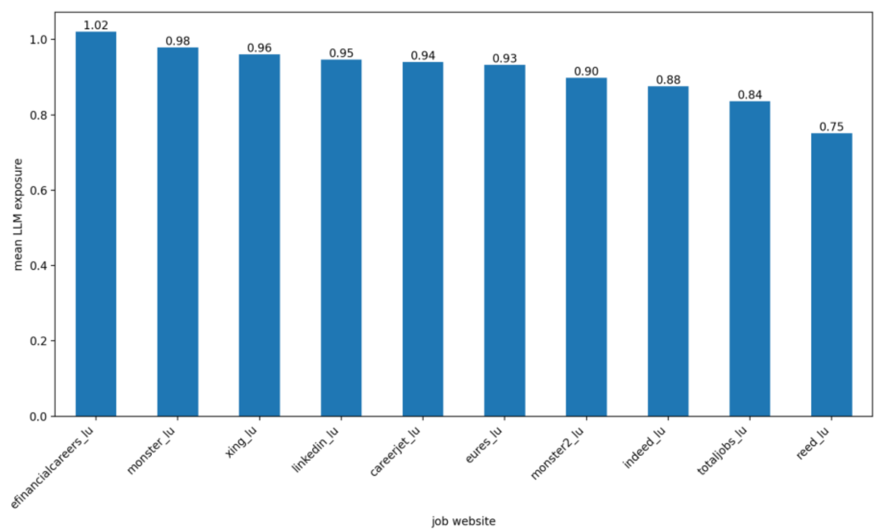


Image 5. Mean LLM exposure per job website in sample



Since source could correlate both with the treatment and the outcome, the relevant control ensures that the "source effect" does not enter β_q estimates, neutralising possible bias. Additionally, inclusion of controls reduces standard errors and narrows down confidence intervals for β_q.

The standard errors are clustered at occupation (SO) level, the same at which LLM exposure is assigned. This ensures that random shocks, common to vacancies within the same occupation, are reflected in the analysis to not underestimate standard errors.

4 Empirical Analysis

4.1 Results

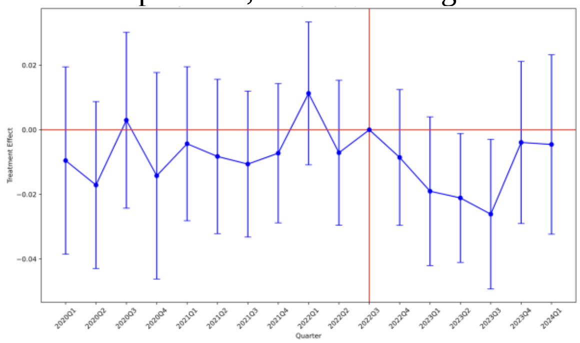
LLM effects on competency demands are identified through the sign of β_q , therefore the primary evidence for any effect is a confidence interval that is entirely above or below 0. The ATE/ACR direction, rather than magnitude, is the primary parameter of focus.

Set out in the tables below are the estimates obtained. Table 4 describes β_q estimates from specification (1), where the treatment $LLM\_exposure_i$ is discrete; Table 5 - from specification (2), where the treatment is continuous. As mentioned in section 3.2, extensive margin specifications have a binary dependent variable indicating the probability that a specific competency occurs in a specific job ad; intensive margin specifications - a continuous dependent variable indicating the number of times competency-related keywords appear in a job ad. The x-axis indicates time, with a vertical bar at 2022Q3, the omitted "baseline" quarter. The y-axis indicates β_q magnitude. The bars show 95% confidence intervals, with standard errors clustered at occupation (SO) level.

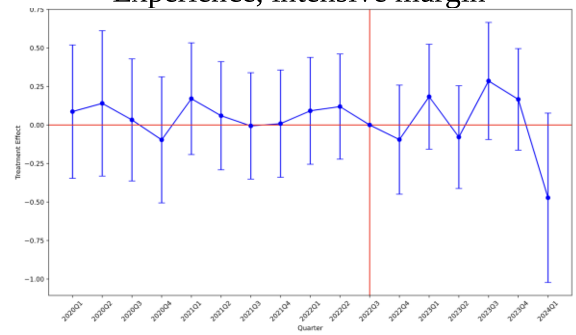
For all specifications except cognitive skill demands, the parallel trends assumption holds in the pre-treatment period, suggesting parallel trends after 2022Q3.

Table 4. Selected ATE Estimates[1]

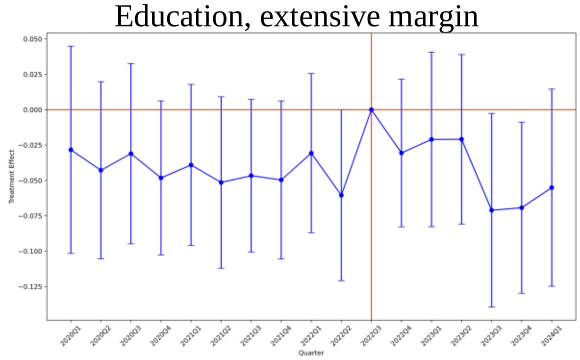
Experience, extensive margin



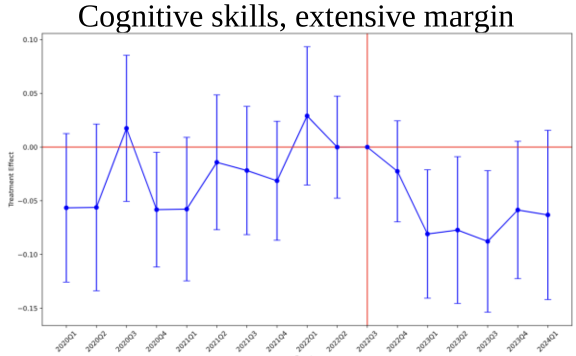
Experience, intensive margin



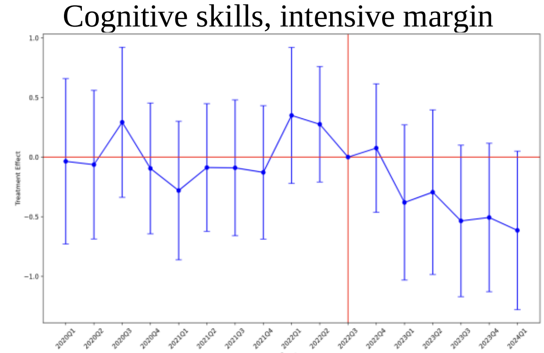
Education, extensive margin



Cognitive skills, extensive margin



Cognitive skills, intensive margin



[1] More details about point estimates and other variables investigated available in Appendix E

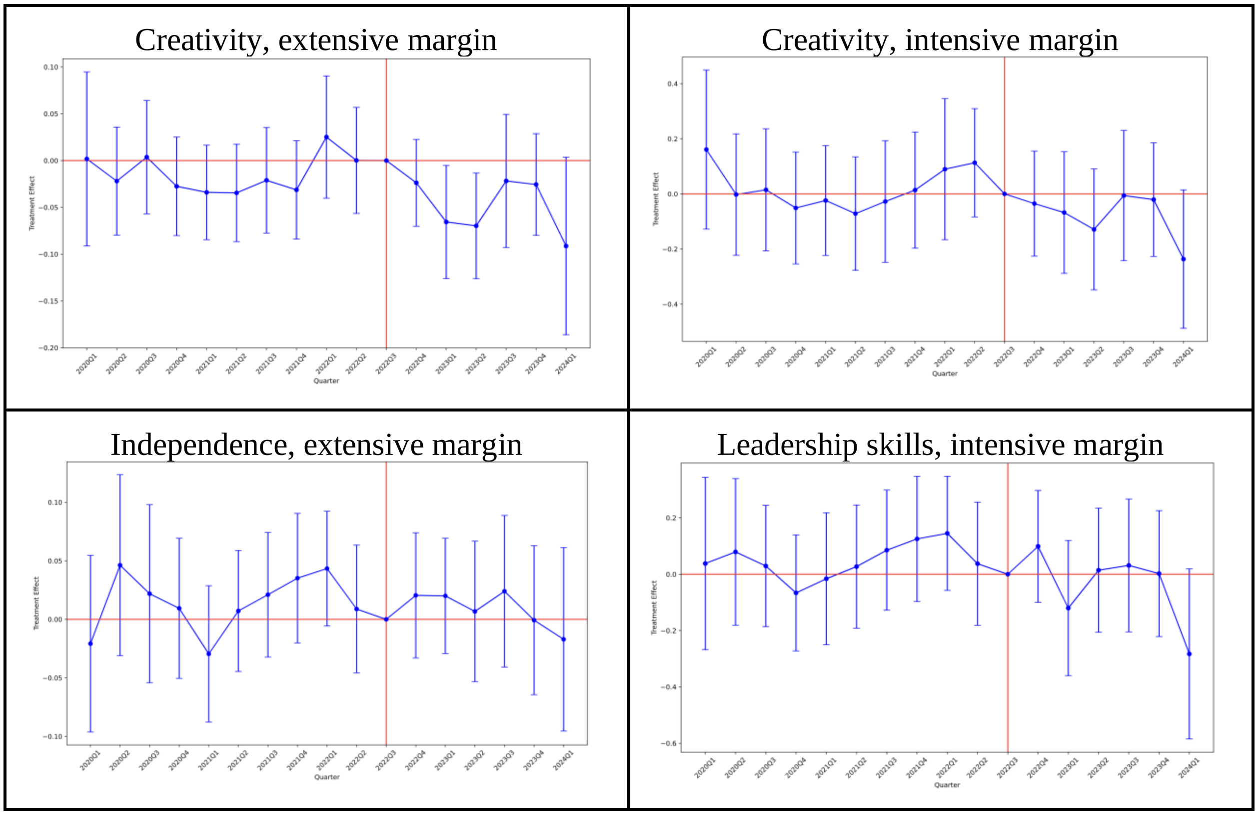
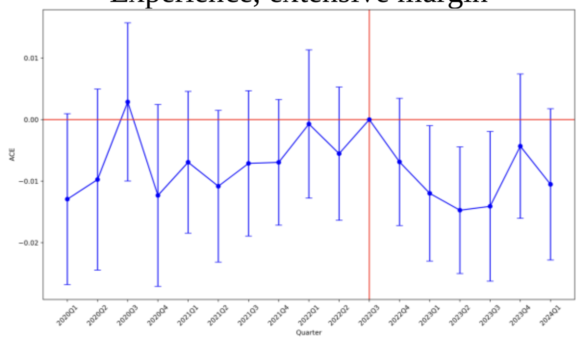
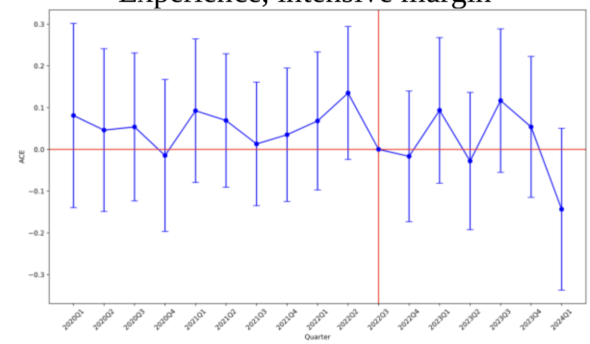

Creativity, extensive margin
Creativity, intensive margin
Independence, extensive margin
Leadership skills, intensive margin
Treatment Effect
Quarter

Table 5. Selected ACR Estimates[2]

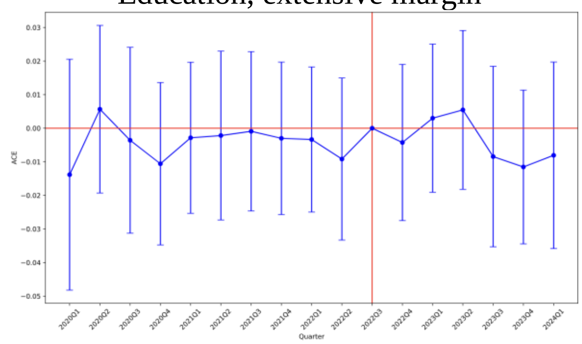
Experience, extensive margin



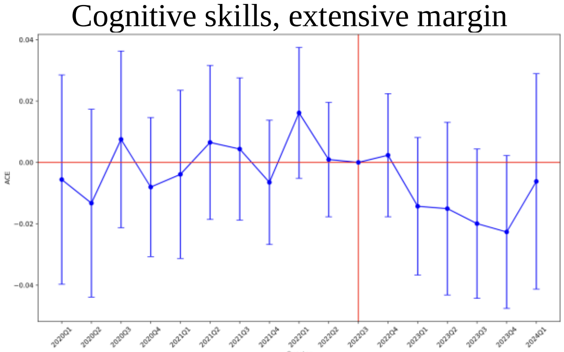
Experience, intensive margin



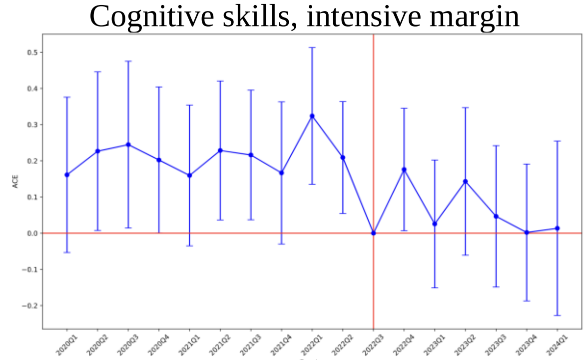
Education, extensive margin



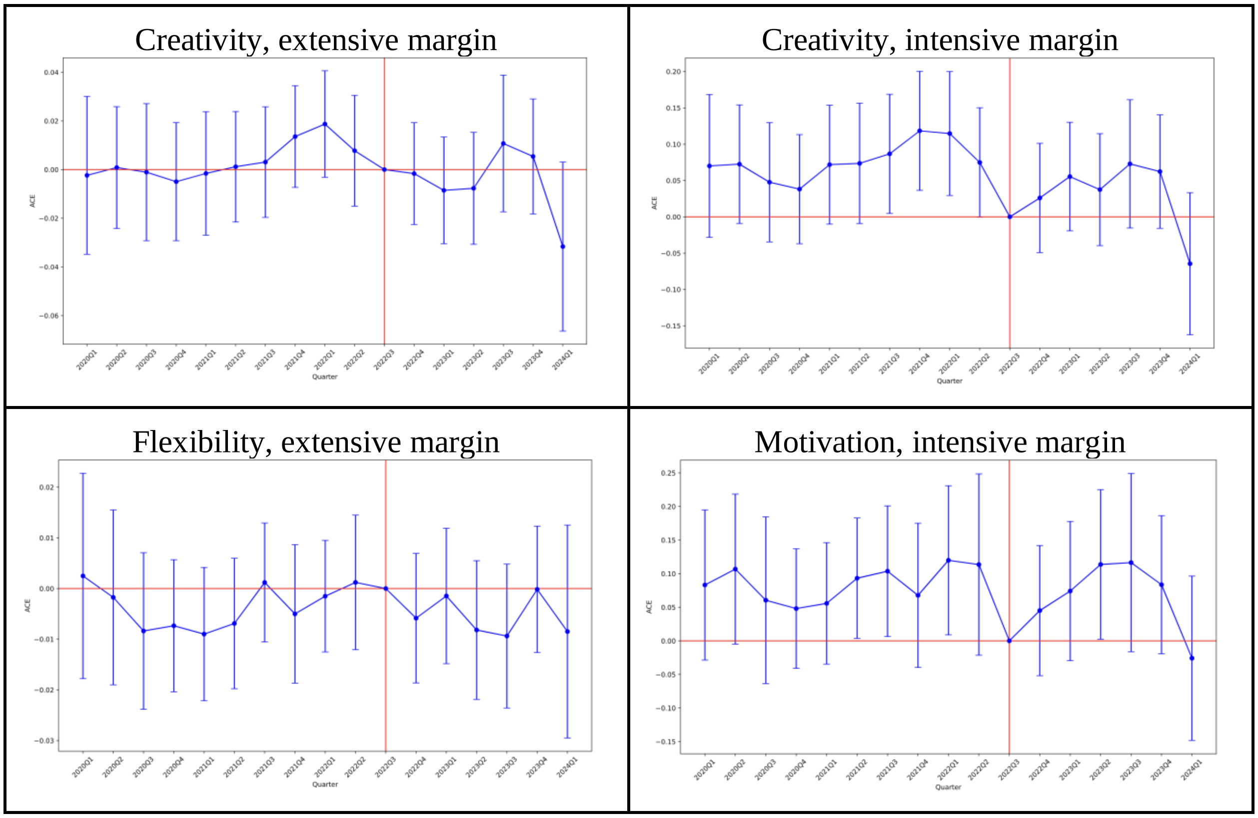
Cognitive skills, extensive margin



Cognitive skills, intensive margin



[2] More details about point estimates and other variables investigated available in Appendix E



Experience. Along the extensive margin, there is a negative ATE/ACR, statistically significant in 2021Q1-2023Q3, but not persistent after. Negative coefficients suggest a negative effect of LLMs on demand for experienced workers. Along the intensive margin (the number of years of experience required), no statistically significant ATE/ACR is identified.

Education. Education is only measured across the extensive margin: whether a vacancy requires any kind of degree. Both the ATE and ACR are negative. The ATE, statistically significant in 2023Q3-2022Q4, is not persistent after, and the ACR is not significant. While negative coefficients indicate that LLMs make education less relevant, the absence of an ACR and mostly insignificant coefficients across specifications suggest no LLM effects on educated labour demand.

Cognitive skills. Estimation results suggest a negative impact of LLMs on cognitive skill demands. The ATE across the extensive margin is the most conclusive, supporting parallel trends and indicating a statistically significant negative effect in 2023Q1-2023Q3. ACR estimates on the intensive margin also support negative LLM effects, despite breached parallel trends. β_q is mostly positive before treatment, indicating that cognitive skill demand grew faster over time for exposed than non-exposed occupations. However, post-shock, this divergence disappears: the ACR is around 0. ATE on the intensive margin and ACR on the extensive margin are negative but insignificant.

Creativity. Estimation shows predominantly no significant ATE or ACR for creativity demands, although there is some evidence for a negative ATE along the extensive margin.

Leadership, independence, flexibility, motivation, time management. In all specifications, leadership, independence, flexibility, motivation and time management skill demands do not seem to be affected by LLMs (full results in Appendix E).

***

The only control variable included is "source" - an indicator for the job board to which a specific posting belongs. It reflects the variation, over time, in the proportion of vacancies scraped from different websites, given the difference in vacancy AI exposure between these websites (see section 3.2). The comparison of R^2 statistics for specifications with and without the "source" control reveals that this control explains some of the additional variation, while leaving β_q largely unchanged (see Appendix F). This supports the inclusion of "source" in above models, and indicates its inclusion is useful for reducing standard errors and thus estimate confidence intervals.

4.2 Discussion

4.2.1 Interpretation

Overall, there is weak evidence for a negative impact of LLMs on experience and cognitive skill demands, and even weaker evidence for a negative impact on education and creativity demands. However, the mostly negative β_q and confidence intervals predominantly in the negative area allow to rule out positive demand effects of LLMs. The remaining competencies seem not to be affected.

Negative ATE/ACR for the presence of experience requirements are consistent with the initial hypothesis of LLMs reducing the value of experience by helping non-experienced workers "bridge the gap" via increased accessibility of relevant knowledge. However, insignificant estimates for years of experience required support the opposite: that experience provides an advantage in terms of behaviours, such as managerial ability, industry awareness or even intuition, which are not AI-replaceable. This contradiction could be resolved if LLMs make experience less relevant for entry-level hires where experience determines knowledge of job technicalities, but still important for more experienced workers who gain other advantages through experience.

Negative ATE for education are consistent with the hypothesis that demand for university-educated labour falls due to LLMs increasing the accessibility of knowledge. However, no effects in ACR estimation suggests that AI might not yet be seen as a substitute for education, as university degrees do not only grant knowledge, but also signal discipline and intellectual ability.

Estimation results for cognitive skill demands suggest they are negatively affected by LLMs, which contradicts the initial hypothesis. However surprising, this negative effect could be primarily driven by reduced demand for information-processing and analysis, rather than problem-solving or decision-making, since LLMs excel at the former but struggle with the latter (Dell'Acqua et al., 2023).

The weak, almost non-identifiable, effect of LLM exposure on creativity demands is inconsistent with expectations: LLMs, primarily known for their creative capabilities, would be expected to strongly substitute creative labour. A possible explanation is that the "office jobs" most represented in the sample interpret creativity in different ways, not necessarily implying text generation abilities or inventiveness.

Finally, as expected, leadership, independence, flexibility and motivation demands are unaffected by LLM exposure, being inherently human and hardly AI-replaceable.

4.2.2 Limitations

The most significant shortcoming of the study is its low power. In fact, low R^2 values[3] (0.002 lowest to 0.118 highest) indicate that most of the variation in the data remains unexplained, which contributes to higher standard errors and less precise estimates (see Appendix F). First, this is due to dataset limitations: a small country, Luxembourg represents a relatively small sample of job adverts available for analysis. Unfortunately, the Luxembourg dataset was the only existing one suitable for this project: freely accessible[4] and spanning sufficient time periods before and after ChatGPT. The number of observations in the sample is further constrained by the fact that ChatGPT was introduced very recently. Additionally, the study power is limited by the issue of multiple testing: the large number of estimated parameters per equation, of dependent variables and of specifications increases the likelihood of finding at least one statistically significant effect purely by chance. In fact, if applying Bonferroni-adjusted alphas (Weisstein, 2024), none of the effects would be statistically significant. The overall implication of low power is the risk of overstating LLM impacts.

Next, there was a notable discrepancy between specifications: the discrete one has materially higher rates of significant ATE/ACR detection than the continuous one. This highlights the relative flexibility of the discrete specification: while the continuous specification imposes a linear relationship between LLM exposure and competency demands, the discrete one only imposes a "cutoff". This suggests a possible non-linear relationship between LLM exposure and competency

[3] Total R^2: proportion of the variation in the dependent variable explained by all the explanatory variables (fixed effects, interaction variables, controls)
[4] The University of Warwick prohibits the use of paid data sources for undergraduate dissertations

demands, therefore experimenting with alternative functional forms of LLM exposure makes a viable next step.

Additionally, in many cases the statistically significant effects were not persistent: β_q often "returned" to levels comparable with pre-treatment periods after 2-3 quarters. This could have occurred due to employers initially overestimating the substitutability of some competencies by AI, or due to different seasonal hiring trends between vacancies. More time periods are needed to confirm or deny these effects. And indeed, insufficient "post-treatment" periods available for researching labour market effects of LLMs pose big challenges for identification (Babina et al., 2022). However, a similar methodology could yield clearer results in the future.

Another reason for inconclusive estimates is noisy data. This can be addressed by including additional controls, such as industry, firm size or job seniority level, extractable from the ad texts with more advanced methods. Moreover, omitted controls that correlate with both skills required in the job (outcome) and its AI exposure (treatment) - for instance, industry - can cause error endogeneity and biased estimates. Overall, insufficient controls represent a major drawback of this study, exacerbating the issue of low power.

Other detection issues may have occurred due to attenuation bias. For example, random errors could be caused by the imperfections in the dictionary-based search used for dependent variable creation. Limited by the specific words included in the dictionaries (Appendix D), this method does not fully capture competency demands, increasing the level of "noise", thus increasing standard errors and reducing estimate precision. Ways to address attenuation bias include focusing on longer (more detailed) vacancy postings to obtain more precise estimates, or normalising the competency scores to the length of the "requirements" section in the posting. Another source of random error and "noise" is the vacancy-SO matching procedure, which does not always generate perfectly accurate matches. This is addressed via matching algorithm improvements.

Another detection difficulty might have occurred because job advert content, as a measurement, does not capture the entirety of competency demands in the labour market. This measurement does not account for the demand for upskilling of already employed professionals, redundancies, promotions, salaries of people with different skillsets or changes in their working hours. Using these concepts to measure labour demand could be an important next step to support or challenge the findings of this study.

Next, LLM exposure scores used to construct the treatment variable are based on subjective human assessment (Felten et al., 2023), moreover, Felten's scores specifically might overestimate the effects of AI (Acemoglu et al., 2022). These scores can also create anticipation bias: the more "observable" the impact of LLMs on a job is, the easier it is to measure exposure. If this anticipation effect boosts exposure scores for affected occupations and reduces exposure scores for non-affected occupations, the effect of interest could be overestimated. This gives reasons to doubt any statistically significant estimates.

Finally, while parallel trends in specification (1) generally hold, the "strong parallel trends" necessary for identification with continuous treatment might not hold universally. "Strong parallel trends" restrict causal response heterogeneity, imposing that different occupations should experience LLM shocks of the same intensity in the same way (Callaway et al., 2024). This is not guaranteed, given the multitude of mechanisms shaping the interaction of AI and labour demands (production function effects, supply effects, demand effects). For instance, lawyers and programmers could face different skill demand dynamics due to the LLM adoption specifics in their respective industries. "Strong parallel trends" cannot be tested, highlighting the need to benchmark continuous treatment estimators against other specifications.

5 Conclusion

Overall, there is weak evidence for a small-scale negative impact of LLMs on experience, education, cognitive skill and creativity demands. There is no evidence for a positive impact of LLMs on any competency demands, and soft skill demands seem not to be subject to any effects.

Where a negative effect was detected, the most plausible explanation for it is a technical substitution effect: LLM functionality substitutes the competencies, reducing the demand for them. This implies possible negative wage and employment effects for educated, experienced, more intellectually able and more creative workers.

An absence of a positive effect allows to rule out LLM complementarity or augmentation, at least currently. LLM-driven reductions in costs of upskilling that increase skill demand either through lower wages or indirectly through firm output growth seem implausible, since they predict opposite demand effects to what is observed, and because they are likely to take place slower than direct substitution. At the same time, the interference of these processes with the substitution effect can contribute to inconclusive estimates.

The absence of visible LLM influence on demands for leadership, independence, flexibility and motivation indicates consistent demand for these skills, independent of technological progress. The workers' value in the labour market depends on developing these skills - perhaps even more so than before, if LLM presence creates less stringent worker requirements in other domains.

Benchmarking against previous research, the identified substitution effects are in line with Hui et al. (2023), who showcase falling wages and employment in exposed occupations. This project extends Hui et al. (2023) findings to show that similar mechanisms can occur outside "writing" occupations. Additionally, the identification of negative or non-present (but not positive) effects of LLMs on educated labour demand signify that LLM adoption implications are different from those of computerization, which increased demand for educated workers (Autor et al., 2003). Next, the identified substitution of cognitive skills by LLMs contradicts Dell'Acqua et al.' (2023) claims about the problems LLMs experience with some advanced cognitive tasks. Finally, the identification of substitution effects among "white-collar" vacancies suggests that labour market implications of LLMs might be different to those of other AI forms, since Acemoglu et al. (2022) demonstrated that AI increases the demand for “white collar job” skills, such as analysis, marketing and finance.

LLM substitution effects could mean a smaller wage premium for more educated, experienced, creative and intellectually able workers. As this is equivalent to a reduction in wage inequalities in more exposed occupations, it rhymes well with Dell’Acqua et al. (2023), who find that LLMs help to close performance gaps between high and low performers.

As for the policy-makers, given the plausible substitution effects, they should acknowledge the fine line between encouraging LLM innovation and implementing relevant regulations that protect or insure workers.

Finally, the substantial ambiguity in the estimation results and the low power of the study make any conclusive judgements on LLM labour market effects premature. More definitive inferences are only possible after addressing the methodological shortcomings and observing more time periods. Additionally, more insights on LLM effects can be gained through investigating skill demands in countries with different labour force composition than the service-oriented Luxembourg, and by considering treatment effect heterogeneity across occupations.

References

Acemoglu, D., Autor, D., Hazell., J., Restrepo, P., 2022. 'Artificial Intelligence and Jobs: Evidence from Online Vacancies', *Journal of Labor Economics, vol. 40, n. S1, pp. 293-340*

Autor, D.H., Levy, F., Murnane, R.J., 2003. 'The Skill Content of Recent Technological Change: An Empirical Exploration', *The Quarterly Journal of Economics, vol. 118, n. 4, pp. 1279-1333*

Babina, T., Fedyk, A., He, A.X., Hodson, J., 2022. 'Artificial Intelligence, Firm Growth, and Product Innovation', *Forthcoming at the Journal of Financial Economics (JFE)*, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3651052

Brynjolfsson, E., Mitchell, T., Rock, D., 2018. 'What Can Machines Learn, and What Does It Mean for Occupations and The Economy?' *AEA Papers and Proceedings, vol. 108, pp. 43-47.*

Callaway, B., Goodman-Bacon, A., Sant'Anna, P.H.C., 2024. 'Difference-in-Differences with a Continuous Treatment', *National Bureau of Economic research Working Paper n. 32117*, Available at NBER: https://www.nber.org/papers/w32117

Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kelogg, K., Rajendran. S., Krayer, L., Candelon, F., Lakhani., K.R., 2023. 'Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality', *Harvard Business School Technology & Operations Mgt. Unit Working Paper n. 24-013*, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321

Elondou, T., Manning, S., Mishkin, P., Rock, D. (OpenAI, OpenResearch, University of Pennsylvania), 2023. 'GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models', *Working Paper*, Available at: https://openai.com/research/gpts-are-gpts

Felten, E., Raj, M., Seamans, R., 2021. 'Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses', *Forthcoming at the Strategic Management Journal*, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3822412

Felten, E., Raj, M., Seamans, R., 2023. 'Occupational Heterogeneity in Exposure to Generative AI', Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4414065

Hui, X., Reshef., O., Zhou, L., 2023. 'The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market', Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4527336
Koulischer, F., Perray, P., Tran, T.T.H., 2022. 'COVID-19 and the Mortgage Market in Luxembourg', *Journal of Risk and Financial Management, vol. 15(3), n. 114*

Marr, B. (Forbes), 2023. 'The Difference Between Generative AI and Traditional AI: An Easy Explanation For Anyone' Available at: https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/ (Accessed on 10 March 2024)

Noy., S., Zhang, W., 2023. 'Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence', *Working Paper*, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375283

OpenAI, 2023. 'GPT-4 Technical Report', Available at arXiv: https://arxiv.org/pdf/2303.08774v4

Otubusen, A., Sleeman, C. (Nesta), 2021. 'Open Jobs Observatory: Assigning online job adverts to occupation groups', Available at: https://www.nesta.org.uk/project-updates/soc-assignment-ojo/ (Accessed on 4 January 2024)

Porter, J. (The Verge), 2023. 'ChatGPT continues to be one of the fastest growing services ever', Available at: https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference (Accessed on 1 April 2024)

Webb, M., 2020. 'The Impact of Artificial Intelligence on the Labor Market', Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3482150

Weisstein, E.W., 2024. 'Bonferroni Correction', *MathWorld - A Wolfram Web Resource,* Available at: https://mathworld.wolfram.com/BonferroniCorrection.html (Accessed on 24 April 2024)