The Effect of Occupation Level AI Exposure on the Gender Gap in Weekly Labour Input: Evidence from Canada

Abstract

Generative artificial intelligence (AI) can perform tasks common to many occupations, yet its early effects are hard to detect in aggregate employment. This paper asks whether the response appears instead within occupations, in the gap between women and men in weekly labour input, and whether it is larger where occupations are more exposed to AI. Using monthly Canadian Labour Force Survey microdata from 2021 to 2025 for the prime-age sample within the paid employee labour force, and a triple difference design that compares women and men within the same occupation and month, this paper finds that higher occupation level AI exposure led to a more negative post-ChatGPT change in the gap. Moving from the least exposed to the most exposed occupation group, the change in the gap in usual weekly hours is 0.803 hours more negative, and the change in the gap in actual weekly hours is 0.956 hours more negative. Separate DID estimates by gender suggest that the pattern is more concentrated among women, especially for usual weekly hours. These estimates point to a distributional margin within occupations where an early labour market response to AI is visible even while aggregate employment shows little response.

Keywords: artificial intelligence; occupation level exposure; gender gap; weekly hours; labour input; triple difference; Canada

JEL Classification: J16; J22; J23; O33; J24

1. Introduction

In November 2022, ChatGPT was released to the public, making generative artificial intelligence (AI) widely available for ordinary work tasks. AI systems can produce text, process information, and support communication, and these activities appear in many occupations. In Canada, occupations classified as highly exposed and less complementary to AI have not experienced a sustained decline in employment (Mehdi and Frenette 2026). If AI is changing work, the early response may therefore appear less in total employment and more in narrower labour input margins within occupations.

Research on gender gaps points to one such margin. Within the same occupational title, women and men can perform different tasks and use different skills (Pető and Reizer 2021; Fana et al. 2023), and hours have long been central to gender differences in labour market outcomes (Goldin 2014; Blau and Kahn 2017). Together, these facts make the gap between women and men in weekly labour input a natural margin for studying early labour market adjustment to AI.

This paper asks whether the post-ChatGPT change in this gap varies with occupation level AI exposure. It uses public monthly microdata from the Statistics Canada Labour Force Survey, covering January 2021 to December 2025. The main analysis focuses on the prime-age sample (age 25-54) within the paid employee labour force. The outcomes are usual weekly hours and actual weekly hours. For unemployed respondents in the paid employee labour force, both outcomes are set to zero, so the outcomes measure weekly labour input rather than only hours among employed workers. The primary exposure measure is the occupation level AI exposure index from Massenkoff and McCrory (2026), and the exposure measure from Eloundou et al. (2024) is used for robustness checks. Both exposure measures are defined at the occupation level and do not reflect the AI use of individual respondents. The empirical

strategy uses a triple difference (DDD) design that compares women and men within the same occupation and month, so shocks common to both, including those from the pandemic recovery, do not enter the comparison.

The main estimates show that the post-ChatGPT change in the weekly labour input gap between women and men, defined as women minus men, is more negative in occupations with higher AI exposure than in occupations with lower AI exposure. Moving from the least exposed to the most exposed occupation group, the change in the gap in usual weekly hours is 0.803 hours more negative, and the change in the gap in actual weekly hours is 0.956 hours more negative. The first estimate is statistically significant at the 1 percent level and the second at the 5 percent level. The paper also estimates separate difference in differences (DID) diagnostics for women and men. For usual weekly hours, these diagnostics suggest that the pattern is more concentrated in the women sample. For actual weekly hours, the estimates are negative in both gender samples and more negative in the women sample.

The paper also reports timing and robustness checks. The event study estimates do not show statistically significant pre-period divergence, and the joint tests do not reject that the pre-period coefficients are jointly zero. The main pattern remains consistent when the primary exposure measure is replaced with the Eloundou measure, when fixed effects and controls are added step by step, when the hours scale is changed, and in a specification that removes the period of the 2025 trade policy changes between the United States and Canada. Placebo tests with earlier fake event timing do not detect the main pattern. Omitting any single occupation cluster does not change the direction of the estimates, and the estimates remain statistically significant under cluster jackknife inference.

This paper makes three contributions. First, it provides evidence on a margin that early work on AI and labour markets has not examined. Mehdi and Frenette (2026) and Humlum and

Vestergaard (2025) study average outcomes, including employment, hours, and earnings, and average outcomes can hide changes in the gap between women and men within occupations. This paper examines this gap directly. To the best of my knowledge, this is the first causal estimate of the effect of occupation level generative AI exposure on the post-ChatGPT change in the gap between women and men in weekly labour input. Second, the results provide new evidence for a task level view of how AI reaches the labour market (Autor et al. 2003; Acemoglu and Restrepo 2019). The design compares women and men within the same occupation, two groups whose task content differs in survey evidence (Pető and Reizer 2021; Fana et al. 2023). A shock that reached women and men in an occupation equally would leave the gap between them unchanged. The finding that the change in this gap varies with exposure therefore reflects responses that differ between women and men within the same occupation. Third, the paper adds evidence from a new technology episode to research on gender gaps in hours. Technological change has altered the relation between hours and the gender earnings gap before (Goldin and Katz 2016). This paper provides evidence on how the hours margin responds during the first years of generative AI.

The remainder of the paper proceeds as follows. Section 2 situates the analysis in the related literature, Section 3 describes the data and measurement, Section 4 presents the empirical strategy, Section 5 reports the results, and Section 6 concludes.

2. Literature Review

Studies using experiments and firm data show that generative AI can improve performance in specific work tasks. Noy and Zhang (2023) show in a writing experiment that ChatGPT reduces completion time and raises output quality, with larger gains for initially weaker writers. Brynjolfsson et al. (2025) find, using customer support data from a firm, that an AI assistant increases the number of issues resolved per hour, with the largest gains among less

experienced workers. Together, they show that generative AI can reduce time input, raise output quality, or increase output per hour in some tasks.

Evidence on average labour market outcomes is less clear than the task evidence. Mehdi and Frenette (2026) find no sustained employment decline in Canadian occupations classified as highly exposed and less complementary to AI, and they find no clear slowdown in employment or weekly earnings growth in more exposed industries. Audoly et al. (2026) find that US job postings declined relatively more in occupations with greater AI exposure, but the decline began before the release of ChatGPT and stabilized after 2023. Humlum and Vestergaard (2025) show that chatbot adoption is widespread in Denmark and report precise nulls for average adoption effects on earnings and recorded hours. These studies do not show a single direction in early average labour market outcomes. Productivity gains are visible in task settings, while average changes in employment, hours, or earnings remain muted.

The task literature explains how these two facts can coexist. Autor et al. (2003) analyze computerization at the level of tasks rather than only at the level of occupations, and they show that computers substitute for routine tasks while complementing tasks involving nonroutine problem solving and communication. Acemoglu and Autor (2011) formalize this logic in a framework that separates skills from tasks. In that framework, technological change can reallocate skills across tasks without requiring labour demand for an entire occupation to move in one direction. Acemoglu and Restrepo (2019) further show that the net effect of automation on labour demand depends on displacement, productivity, capital accumulation, and the creation of new tasks. Generative AI can therefore create both substitution and complementarity across tasks. When these forces operate at the same time, occupational or market averages may show little net change even though task level outcomes have begun to adjust.

Evidence on tasks and gender gaps shows that women and men may perform different tasks in the same or closely comparable work settings. Fana et al. (2023) use task data collected from workers in France and find systematic task differences between women and men within the same job. Whelan et al. (2026) provide related evidence for digital tasks. They find that women are less likely than men to perform advanced digital tasks, and this difference remains after controlling for occupation, industry, and firm characteristics. Stinebrickner et al. (2026) use data on individual time allocation and show that women and men allocate working time differently across tasks. Women spend more time on people tasks, while men spend more time on information tasks that require higher skill. Pető and Reizer (2021) add evidence on skill use. They show that women use cognitive skills less than men within the same country and occupation, conditional on cognitive test scores. These studies show observable differences in actual task content and skill use between women and men. They motivate the paper's focus on outcome gaps between women and men.

Hours and time availability are classic outcome margins in research on gender gaps. Goldin (2014) argues that long hours, specific time availability, and convex returns to hours help sustain gender gaps. Goldin and Katz (2016) show in pharmacy that greater worker substitutability and a more linear relation between pay and hours are associated with a much smaller gender gap. Blau and Kahn (2017) also emphasize that shorter hours remain important for understanding gender gaps after the declining explanatory role of traditional human capital variables. This paper adopts this hours margin and examines the gap in weekly labour input between women and men.

Generative AI changes performance in specific work tasks, and actual task content differs between women and men within the same occupation. Together, they imply that the post-ChatGPT change in the gap between women and men in weekly labour input may vary with occupational AI exposure. This is the question of the paper. Answering it requires separating

AI exposure from concurrent developments in the labour market. In Canada, the post-ChatGPT period coincided with the later stages of the pandemic recovery and rapid population growth driven by immigration (Mehdi and Frenette 2026). A comparison between more exposed and less exposed occupations therefore cannot distinguish the consequences of AI exposure from differences in occupational recovery paths. Within the same occupation and month, however, women and men face the same occupational conditions, so a shock to the occupation as a whole leaves the gap between them unchanged. The paper therefore compares the change in this gap across occupations with different AI exposure. Changes that affect women in all occupations alike move every occupation's gap by a similar amount and do not enter this comparison. The gap itself is measured within occupations, which is the level at which the task differences have been documented. The Labour Force Survey supports this design with monthly public microdata. The survey records usual and actual weekly hours for employed respondents and retains the last job occupation for unemployed respondents, although individual tasks are not observed. The evidence above motivates the comparison but does not restrict the direction of the change, which remains an empirical question.

3. Data and Sample

This paper uses the Public Use Microdata File of the Statistics Canada Labour Force Survey, hereafter the LFS, as its source of individual level labour market microdata. The LFS covers the civilian household population aged 15 and over in the ten Canadian provinces. The microdata file is a repeated cross section, and each observation is one person in one survey month.[1] The LFS provides the core variables used in this paper, including gender, age group, education, immigrant status, province, survey month, occupation, class of worker, labour

[1] The microdata file does not link the same respondent across months, so each person month is treated as a separate observation.

force status, usual and actual weekly hours, and the LFS final survey weights. The analysis window runs from January 2021 to December 2025 (2021m01 to 2025m12).

This paper merges the LFS person month records with occupation level AI exposure data. Both exposure measures are constructed from task level information, so occupation level exposure reflects the tasks that AI can perform or assist within each occupation. The Anthropic observed exposure measure (Massenkoff and McCrory 2026) combines theoretical task capability with observed work-related Claude usage. The Eloundou beta measure (Eloundou et al. 2024) captures task-level exposure based on the capabilities of large language models rather than observed use.[2] In this early period, actual AI use covers part of the tasks that are feasible for AI (Massenkoff and McCrory 2026), so the two measures overlap but are not identical. This paper therefore takes the Anthropic measure as the primary exposure measure and uses the Eloundou measure in the appendix for robustness.[3] The original exposure scores are defined on O*NET SOC occupations, and this paper maps them into the Canadian NOC through public crosswalks and aggregates them to 43 NOC occupation groups, hereafter NOC43 occupation groups.[4] Each exposure measure is separately normalized to the [0,1] interval across the 43 occupation groups using a min max normalization.[5]

[2] For the Anthropic measure, theoretically feasible tasks enter the index only when they satisfy a minimum usage threshold in work-related Claude traffic. Automation receives full weight, augmentation receives one-half weight, and tasks are combined using estimated task-time shares. The Eloundou measure gives a value of one to tasks for which an LLM alone could reduce completion time by at least half, one-half when complementary software is needed, and zero otherwise.

[3] The usage information in the Anthropic measure comes from aggregate worldwide AI usage, not from the respondents in this sample. Neither measure reflects the AI adoption or use of the respondents, and the exposure values do not depend on any labour market outcome reported in the LFS. The exposure values are also fixed over the analysis window.

[4] The detailed mapping moves from O*NET SOC scores to SOC level means, then to NOC 2016 through an unweighted mean, then to NOC 2021 through a share weighted concordance, and finally to unweighted means for the 43 NOC occupation groups. Mapping US O*NET occupation measures into the Canadian NOC classification by averaging to the occupation level follows Beland et al. (2022).

[5] For each measure, normalization maps the observed minimum across the 43 occupation groups to zero and the observed maximum to one.

The main sample is the prime-age sample (age 25-54). This age range covers the core working-age population. The sample excludes ages 15 to 24 and ages 55 to 64, where participation and hours also reflect schooling and retirement margins (Blundell et al. 2013). Within the prime-age range, the analysis is further restricted to the paid employee labour force, the main analysis domain of this paper. This domain includes labour force respondents whose current job, or whose last job in the case of the unemployed, is in public or private paid employment.[6] Respondents outside the labour force do not enter the main analysis domain. Self-employed workers and unpaid family workers are also outside this domain.

The sample must also link to a NOC43 occupation group and its occupation exposure through the LFS occupation codes. The baseline consists of all prime-age person months from 2021m01 to 2025m12 that are recorded as women or men and carry a valid survey weight, a total of 2,796,069 person months. Of this baseline, 2,492,500 person months (89.1 percent) link to occupation exposure. The final main sample is the 2,133,148 person months (76.3 percent of the baseline) that fall in the paid employee labour force domain. Both exposure measures are available for all 43 NOC43 occupation groups in the final main sample. The appendix also uses an all-working-age sample (age 15-64). This sample follows the same construction as the prime-age sample and only widens the age range.

This paper uses two hours outcomes, usual weekly hours and actual weekly hours. Both outcomes measure total weekly hours across all jobs. For employees, usual weekly hours are normal paid or contract hours and do not include overtime. Actual weekly hours are the hours actually worked in the reference week. They include paid and unpaid hours and reflect temporary events such as illness, vacation, holidays, and overtime. Each person month carries

[6] For respondents who are not currently employed but worked within the previous 12 months, the LFS retains the class of worker and occupation of the last job. The survey therefore identifies whether the previous job was in public or private paid employment.

one NOC43 occupation link that assigns exposure. For employed respondents, this link is the current main job; for unemployed respondents, it comes from the last job.

The main hours outcomes are constructed within the paid employee labour force domain. For employed respondents, the outcomes equal their reported hours. Unemployed respondents in this domain are assigned zero hours. The two outcomes therefore combine hours changes among the employed with changes in the employment and unemployment composition of the sample, and they measure weekly labour input rather than a pure intensive margin among employed workers. This construction follows labour supply research that treats work status and hours as related components of total labour supply.[7]

Table 1 reports descriptive statistics for the main analysis sample. The sample covers 43 occupation groups and 10 provinces. In Panel D, exposure runs from 0.068 at the 25th percentile to 0.627 at the 75th percentile in the full sample. The empirical strategy draws on this variation across occupation groups. Panel D also shows that exposure is higher for women than for men across the distribution. Exposure is assigned at the occupation level, so women and men in the same occupation group carry the same exposure value. The gender differences in Panel D therefore reflect occupational composition. The occupation-by-female fixed effects absorb the time invariant outcome differences between women and men within each occupation group, so the estimates do not rely on these differences in occupational composition. Appendix Table A.1 summarizes the Anthropic and Eloundou exposure measures by broad occupation group.

This paper does not estimate an aggregate effect for higher exposure occupations. The core question is whether the weekly labour input gap between women and men changes more in

[7] Labour supply research often measures hours including the nonemployed and treats participation and hours as joint margins (Blau and Kahn 2007; Blundell et al. 2013). Measuring hours conditional on employment instead captures a narrower object (Rothstein 2010).

occupations with higher AI exposure after ChatGPT. This question calls for a gender contrast within the same occupation and month.

4. Empirical Strategy

4.1 Main DDD Design

The main empirical strategy is a triple difference (DDD) design. It estimates how the post-ChatGPT change in the weekly labour input gap between women and men varies with occupation level AI exposure. The design compares weekly hours across occupations with different AI exposure, across months before and after ChatGPT, and between women and men. Equation 1 is the main DDD equation, and the term of interest is the triple interaction $Exposure_o \times Post_t \times Female_i$.

$$Y_{iot} = \beta\,(Exposure_o \times Post_t \times Female_i) + \alpha_{o\times t} + \lambda_{p\times t} + \mu_{o\times Female} + \nu_{t\times Female} + \rho_{p\times Female} + X_i'\Gamma + \varepsilon_{iot}. \quad (1)$$

In Equation 1, i indexes respondents, o indexes NOC43 occupation groups, and t indexes survey months. The outcome Y_{iot} is usual weekly hours or actual weekly hours. In the fixed effect terms, p indexes the province of respondent i. Equation 1 is estimated separately for each outcome, which matches the column structure of Table 2. $Exposure_o$ is a continuous AI exposure measure at the level of the 43 NOC occupation groups, normalized to [0,1]. ChatGPT was released in November 2022, so the event month is 2022m11. The pre-period is 2021m01 to 2022m10, and the post-period is 2022m11 to 2025m12. The post indicator $Post_t$ equals one in the post-period. $Female_i$ indicates women. The gender contrast is taken within the same occupation and month. Within an occupation group, women and men share the same measured exposure. The gap in weekly labour input is defined as women minus men. β measures how the post-ChatGPT change in the gap in weekly labour input between women and men varies with occupation level AI exposure. It equals the fitted difference in that

change between the most exposed and the least exposed occupation groups on the normalized exposure scale. X_i is a vector of individual level demographic controls, and Γ is the corresponding coefficient vector. ε_{iot} is an observation level error term.

Equation 1 includes five sets of fixed effects. The occupation-by-month fixed effects $\alpha_{o\times t}$ absorb shocks that women and men in the same occupation and month face together. The province-by-month fixed effects $\lambda_{p\times t}$ absorb shocks common to women and men within the same province and month. The occupation-by-female fixed effects $\mu_{o\times Female}$ absorb time invariant gender differences within each occupation. The month-by-female fixed effects $\nu_{t\times Female}$ absorb gender specific monthly movements that are common to all occupations. The province-by-female fixed effects $\rho_{p\times Female}$ absorb time invariant differences in the gender gap across provinces. Because exposure is defined at the occupation level and the post indicator varies at the month level, the lower order interactions $Exposure_o \times Post_t$, $Exposure_o \times Female_i$, and $Post_t \times Female_i$ are absorbed by the occupation-by-month, occupation-by-female, and month-by-female fixed effects, respectively. The identifying assumption is that, in the absence of ChatGPT, the weekly labour input gap between women and men would have evolved similarly across occupations with different AI exposure. Under the identifying conditions for the continuous exposure specification,[8] β identifies the slope of the average causal exposure-response function, which is modeled linearly over the observed exposure range. The event study and placebo tests in Section 5 report pre-period evidence consistent with the parallel-trends assumption.

[8] Callaway, Goodman-Bacon, and Sant'Anna (2024) show that causal comparisons across levels of a continuous treatment rely on a strong parallel trends condition. In the present setting, conditional on the fixed effects and covariates in Equation 1, this condition requires that, for a given exposure level, the average counterfactual evolution of the gender gap be the same across occupations with different realized exposure levels.

X_i includes demographic controls for age group, education, and immigrant status. All regressions use the LFS final weights. Standard errors are clustered at the occupation level. The outcomes are in levels of weekly hours, so the coefficients are measured in weekly hours.

4.2 Gender Sample DID

Equation 2 is a source diagnostic for the main DDD result, estimated as a difference in differences (DID) specification separately in the women sample and in the men sample. The term of interest is the interaction $Exposure_o \times Post_t$.

$$Y_{iot} = \beta_g(Exposure_o \times Post_t) + \alpha_o^g + \lambda_p^g + \tau_t^g + X_i'\Gamma_g + \varepsilon_{iot}^g, \quad i \in g,\ g \in \{\text{Women, Men}\}. \tag{2}$$

In Equation 2, g indexes the gender sample and takes the values women or men, and $i \in g$ means that the observation belongs to that sample. β_g measures how the post-ChatGPT change in weekly labour input varies with occupation level AI exposure within gender sample g. It equals the difference in that change between the most exposed and the least exposed occupation groups on the normalized exposure scale. In Table 3, the Women DID columns report β_{Women}, and the Men DID columns report β_{Men}. The terms α_o^g, λ_p^g, and τ_t^g are additive occupation, province, and month fixed effects within gender sample g. Γ_g is the coefficient vector on the demographic controls within gender sample g, and ε_{iot}^g is the corresponding error term. The main effects of $Exposure_o$ and $Post_t$ are absorbed by the occupation and month fixed effects, respectively, so Equation 2 includes only the interaction term.

Equation 2 shares with Equation 1 the outcomes, the exposure measure, the post indicator, the event month, the demographic controls, the weights, and the occupation level clustering. Its fixed effects structure is additive and differs from the saturated structure of Equation 1.

The function of Equation 2 is to indicate whether the source pattern behind the DDD gap comes from the women side, the men side, or both sides.

5. Results

5.1 Main Results

Table 2 reports the main DDD estimates for usual weekly hours and actual weekly hours in the prime-age sample, age 25-54, within the paid employee labour force. The coefficient on AI exposure × post-ChatGPT × female estimates the effect of occupation level AI exposure on the post-ChatGPT change in the gap in weekly labour input between women and men. Columns (1) and (2) use usual weekly hours, and columns (3) and (4) use actual weekly hours. Columns (2) and (4) add demographic controls for age group, education, and immigrant status. Columns (1) and (3) show that the estimates are similar without these controls. Column (2) reports -0.803 for usual weekly hours, and the estimate is statistically significant at the 1 percent level. Column (4) reports -0.956 for actual weekly hours, and the estimate is statistically significant at the 5 percent level. Within the prime-age sample and the paid employee labour force, the post-ChatGPT change in the weekly labour input gap, defined as women minus men, is more negative in occupations with higher AI exposure than in occupations with lower AI exposure. The column (2) estimate implies that the post-ChatGPT change in the gap in usual weekly hours is 0.803 hours more negative in the most exposed occupation group than in the least exposed occupation group. The column (4) estimate implies that the post-ChatGPT change in the gap in actual weekly hours is 0.956 hours more negative in the most exposed occupation group than in the least exposed occupation group.

Figure 1 is a binned DDD event study used as a pre-period timing check, and Appendix Table A.5 reports the corresponding joint tests for the pre-period binned coefficients. The Figure 1

event study estimates do not show statistically significant pre-period divergence. The Table A.5 p-values are 0.641 for usual weekly hours and 0.895 for actual weekly hours, and these tests fail to reject that the pre-period binned coefficients are jointly zero. For both outcomes, the negative point estimates in the first post interval are smaller in magnitude than the negative estimates in several later intervals, and the subsequent estimates follow a non-monotonic path.

Table 3 reports gender sample DID estimates for the women sample and the men sample. These estimates show whether the pattern in Table 2 appears in the women sample, the men sample, or both.[9] Within each gender sample, a negative estimate means that the post-ChatGPT change in weekly hours is more negative in occupations with higher AI exposure. For usual weekly hours, the women estimate is -0.630 and is statistically significant at the 1 percent level. The men estimate is -0.103 and is not statistically significant. For actual weekly hours, the women estimate is -1.339, the men estimate is -0.641, and both are statistically significant at the 1 percent level. Taken together, the diagnostics suggest that the usual hours pattern is more concentrated in the women sample. The actual hours pattern appears in both samples, with a more negative estimate in the women sample. Table 3 thus complements Table 2 by showing the direction of change within each gender sample.

5.2 Robustness

Appendix Table A.2 reestimates the Table 2 specification, replacing the Anthropic exposure measure with the Eloundou exposure measure. The Eloundou measure captures theoretical capability rather than observed usage, so this check varies the type of exposure measurement rather than only its source. The estimates are -1.362 for usual weekly hours and -1.840 for

[9] These gender sample DID estimates are diagnostic and need not mechanically equal the pooled DDD estimate in Table 2 because the specifications differ.

actual weekly hours, and both are statistically significant at the 1 percent level. These estimates are consistent with Table 2 for both weekly labour input outcomes.

Appendix Table A.3 examines whether the results change as fixed effects and controls are added step by step. All columns include occupation-by-month, occupation-by-female, and month-by-female fixed effects. These fixed effects absorb the lower order interactions, so each column estimates only the DDD coefficient. Columns (2) and (6) add province-by-month fixed effects. Columns (3) and (7) add province-by-female fixed effects. Columns (4) and (8) add demographic controls. For usual weekly hours, the four estimates are -0.808, -0.794, -0.797, and -0.803, and all are statistically significant at the 1 percent level. For actual weekly hours, the four estimates are -0.990, -0.967, -0.967, and -0.956, and all are statistically significant at the 5 percent level. The direction and statistical significance are stable across the fixed effects ladder.

Appendix Table A.6 checks whether the results depend on the scale used for weekly hours. The log(1 + hours) estimates are -0.047 for usual weekly hours and -0.062 for actual weekly hours. The inverse hyperbolic sine (IHS) estimates are -0.053 for usual weekly hours and -0.070 for actual weekly hours. The usual hours estimates are statistically significant at the 1 percent level, and the actual hours estimates are statistically significant at the 10 percent level. The direction is consistent with Table 2.

Appendix Table A.4 compares the all-working-age (age 15-64) sample, the prime-age sample, and the all-working-age sample excluding the prime-age sample (age 15-24 and age 55-64). All six columns use the same DDD specification with demographic controls. Columns (1) and (2) use the all-working-age sample; the estimates are -0.385 for usual weekly hours and -0.494 for actual weekly hours, and both are statistically significant at the 5 percent level. Columns (3) and (4) use the prime-age sample and reproduce columns (2) and (4) of Table 2.

Columns (5) and (6) use the all-working-age sample excluding the prime-age sample; the estimates are 0.389 for usual weekly hours and 0.136 for actual weekly hours, and neither estimate is statistically significant. The all-working-age estimates are consistent with Table 2 in direction and statistical significance, but their magnitudes are smaller than the prime-age sample estimates. In the excluded age groups, the paper does not detect the exposure related pattern. Figure A.1 provides the timing diagnostic for the all-working-age estimates in columns (1) and (2) of Table A.4.

Appendix Table A.7 reports two placebo timing tests that ask whether the main pattern appears under fake event timing before the actual post-ChatGPT period. The first placebo test sets the fake event month to 2022m05, about six months before the actual post-ChatGPT start; its reference window is 2021m11 through 2022m04, and its placebo post window is 2022m05 through 2022m10. The second placebo test sets the fake event month to 2021m11, about one year before the actual post-ChatGPT start; its reference window is 2021m01 through 2021m10, and its placebo post window is 2021m11 through 2022m10. Neither placebo window produces estimates that are statistically significant at the 10 percent level. The placebo tests do not detect the main pattern under earlier fake event timing.

Appendix Table A.8 addresses the concern that 2025 trade policy changes between the United States and Canada could drive the main estimates. It restricts the sample window to 2021m01 through 2024m12. The estimates are -0.856 for usual weekly hours and -1.188 for actual weekly hours, and both are statistically significant at the 5 percent level. When the window ends in 2024m12, the estimates remain negative, statistically significant, and consistent with Table 2. The main pattern therefore does not depend on the 2025 months.

Appendix Table A.9 examines cluster influence and CV3J inference following MacKinnon et al. (2023). It covers four specifications. These are the Table 2 Anthropic exposure

specification and the Appendix Table A.2 Eloundou exposure specification, each for usual weekly hours and actual weekly hours. These specifications cluster standard errors at the occupation level with 43 occupation clusters. Panel A checks whether the results depend on any single occupation cluster. Panel B checks whether inference remains stable under CV3J cluster jackknife inference relative to conventional CV1 inference. Panel A omits one occupation cluster at a time and reestimates each specification. It reports the range of estimates across the 43 reestimations. These ranges remain negative for all four specifications, and the same sign counts are 43 out of 43. The full sample estimates lie within these ranges. Omitting any single occupation cluster therefore does not change the direction of the estimates. Panel B keeps the same sample, fixed effects, demographic controls, weights, and clustering at the occupation level. It compares conventional CV1 p-values with CV3J p-values for each specification. Although CV3J is a more conservative cluster inference method, all four specifications remain statistically significant at least at the 5 percent level under CV3J. Together, Panel A and Panel B indicate that the Anthropic and Eloundou estimates are not driven by any single occupation cluster and remain statistically significant under CV3J inference.

Appendix Table A.10 examines whether changes in observed workforce composition contribute to the main DDD estimates. Post-period observations are reweighted separately for women and men so that each gender's joint distribution of age group, education, and immigrant status matches its pre-period distribution. The reweighted estimates are -0.846 for usual weekly hours and -1.002 for actual weekly hours, close to the corresponding Table 2 estimates of -0.803 and -0.956. This similarity indicates that changes in the observed age, education, and immigrant-status composition of the female and male samples do not appear to drive the main results.

6. Conclusion

This paper asked whether the early labour market response to generative AI appears within occupations rather than in aggregate employment. Using Canadian data for the prime-age sample within the paid employee labour force and a triple difference design, this paper finds that the post-ChatGPT change in the gap between women and men in weekly labour input is more negative in occupations more exposed to AI, and the separate gender diagnostics suggest that this pattern is more concentrated among women, most clearly in usual weekly hours. Because the comparison holds the occupation fixed, the change reflects responses that differ between women and men, two groups whose task content differs in survey evidence. Hours have long been a central margin for gender gaps. The evidence here shows the same margin moving in the first years of generative AI. A natural next step is to link task level exposure to the tasks that women and men actually perform, and to test whether these task differences drive the change in the gap. As AI diffuses, its earliest labour market effects may appear within occupations, on margins that aggregate figures do not show.

References

Acemoglu, Daron, and David Autor. 2011. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, vol. 4B, ed. Orley Ashenfelter and David Card, 1043–1171. Amsterdam: Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5.

Acemoglu, Daron, and Pascual Restrepo. 2019. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives* 33 (2): 3–30. https://doi.org/10.1257/jep.33.2.3.

Audoly, Richard, Miles Guerin, and Giorgio Topa. 2026. Do job postings show early labor-market effects of AI? *Liberty Street Economics*, Federal Reserve Bank of New York, May 14. https://doi.org/10.59576/lse.20260514.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics* 118 (4): 1279–1333. https://doi.org/10.1162/003355303322552801.

Beland, Louis-Philippe, Abel Brodeur, Derek Mikola, and Taylor Wright. 2022. The short-term economic consequences of COVID-19: Occupation tasks and mental health in Canada. *Canadian Journal of Economics/Revue canadienne d'économique* 55 (S1): 214–247. https://doi.org/10.1111/caje.12543.

Blau, Francine D., and Lawrence M. Kahn. 2007. Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor Economics* 25 (3): 393–438. https://doi.org/10.1086/513416.

Blau, Francine D., and Lawrence M. Kahn. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature* 55 (3): 789–865. https://doi.org/10.1257/jel.20160995.

Blundell, Richard, Antoine Bozio, and Guy Laroque. 2013. Extensive and intensive margins of labour supply: Work and working hours in the US, the UK and France. *Fiscal Studies* 34 (1): 1–29. https://doi.org/10.1111/j.1475-5890.2013.00175.x.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. Generative AI at work. *The Quarterly Journal of Economics* 140 (2): 889–942. https://doi.org/10.1093/qje/qjae044.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024. Difference-in-differences with a continuous treatment. NBER Working Paper 32117. https://doi.org/10.3386/w32117.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. GPTs are GPTs: Labor market impact potential of LLMs. *Science* 384 (6702): 1306–1308. https://doi.org/10.1126/science.adj0998.

Fana, Marta, Davide Villani, and Martina Bisello. 2023. Gender gaps in power and control within jobs. *Socio-Economic Review* 21 (3): 1343–1367. https://doi.org/10.1093/ser/mwac062.

Goldin, Claudia. 2014. A grand gender convergence: Its last chapter. *American Economic Review* 104 (4): 1091–1119. https://doi.org/10.1257/aer.104.4.1091.

Goldin, Claudia, and Lawrence F. Katz. 2016. A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics* 34 (3): 705–746. https://doi.org/10.1086/685505.

Humlum, Anders, and Emilie Vestergaard. 2025. Still waters, rapid currents: Early labor market transformation under generative AI. NBER Working Paper 33777, revised March 2026. https://doi.org/10.3386/w33777.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb. 2023. Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust. *The Stata Journal* 23 (4): 942–982. https://doi.org/10.1177/1536867X231212433.

Massenkoff, Maxim, and Peter McCrory. 2026. Labor market impacts of AI: A new measure and early evidence. Anthropic, March 5. https://www.anthropic.com/research/labor-market-impacts.

Mehdi, Tahsin, and Marc Frenette. 2026. Canadian employment trends in the era of generative artificial intelligence: Early evidence. *Economic and Social Reports* 6 (1). Statistics Canada, Catalogue no. 36-28-0001. https://doi.org/10.25318/36280001202600100003-eng.

Noy, Shakked, and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381 (6654): 187–192. https://doi.org/10.1126/science.adh2586.

Pető, Rita, and Balázs Reizer. 2021. Gender differences in the skill content of jobs. *Journal of Population Economics* 34 (3): 825–864. https://doi.org/10.1007/s00148-021-00825-6.

Rothstein, Jesse. 2010. Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. *American Economic Journal: Economic Policy* 2 (1): 177–208. https://doi.org/10.1257/pol.2.1.177.

Stinebrickner, Todd, Ralph Stinebrickner, and Paul Sullivan. 2026. Job tasks, task-specific work experience, and the gender wage gap. *Journal of Human Capital* 20 (1): 1–34. https://doi.org/10.1086/738042.

Whelan, Adele, Luke Brosnan, and Seamus McGuinness. 2026. The gender gap in digital skills at work. IZA Discussion Paper 18675.

Tables and Figures

Table 1: Descriptive Statistics

Statistic	Women	Men	Full sample
Panel A. Sample			
Observations, full window (n)	1,070,651	1,062,497	2,133,148
Observations, pre-period (n)	354,673	348,544	703,217
Observations, post-ChatGPT (n)	715,978	713,953	1,429,931
Occupation groups (n)	43	43	43
Provinces (n)	10	10	10
Panel B. Weekly Labour Input			
Usual hours (hours)	35.2	38.4	36.8
Actual hours (hours)	30.4	35.7	33.1
Unemployment share (%)	3.1%	3.9%	3.5%
Panel C. Demographics			
Age 25-29 (%)	18.1%	18.7%	18.4%
Age 30-34 (%)	18.2%	19.2%	18.7%
Age 35-39 (%)	16.7%	17.4%	17.1%
Age 40-44 (%)	16.7%	16.4%	16.6%
Age 45-49 (%)	15.1%	14.1%	14.6%
Age 50-54 (%)	15.2%	14.2%	14.7%
Postsecondary (%)	83.8%	74.8%	79.2%
Immigrant (%)	29.1%	27.3%	28.2%
Panel D. AI Exposure ([0,1])			
Mean	0.395	0.302	0.348
p25	0.145	0.068	0.068
p50	0.363	0.193	0.250
p75	0.627	0.456	0.627

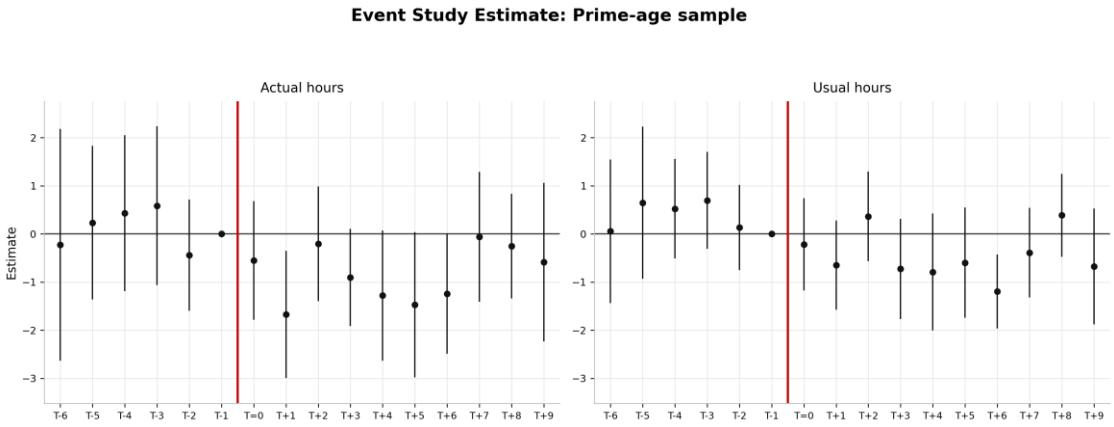
Note: This table reports descriptive statistics for prime-age (age 25-54) LFS respondents in the paid employee labour force. Panel A reports unweighted support counts. Panels B, C, and D report weighted statistics computed over the full window, 2021m01 to 2025m12, with LFS final weights. Weekly hours are measured in hours per week, and unemployed respondents in this sample are assigned zero hours. The unemployment share is measured within this sample. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1].

Table 2: Main DDD Estimates

	(1)	(2)	(3)	(4)
	Usual weekly hours	Usual weekly hours	Actual weekly hours	Actual weekly hours
AI exposure × post-ChatGPT × female	-0.797***	-0.803***	-0.967**	-0.956**
	(0.292)	(0.293)	(0.416)	(0.418)
Occupation × month FE	Yes	Yes	Yes	Yes
Province × month FE	Yes	Yes	Yes	Yes
Female interacted FE	Yes	Yes	Yes	Yes
Demographic controls	No	Yes	No	Yes
Observations	2,133,148	2,133,148	2,133,148	2,133,148

Note: This table reports DDD estimates of how the post-ChatGPT change in the gap in weekly hours between women and men varies with AI exposure. The gap is defined as women minus men. The outcomes are usual weekly hours in columns (1) and (2), and actual weekly hours in columns (3) and (4). The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2025m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 onward. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. Columns (2) and (4) add demographic controls for age group, education, and immigrant status. All columns use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Event-Study Figure



Note: This figure plots event time estimates for the DDD coefficient in the Table 2 specification. The left panel shows estimates for actual weekly hours, and the right panel shows estimates for usual weekly hours. The horizontal axis shows event time bins, and the vertical axis shows estimated coefficients in weekly hours. The event window runs from 2021m01 to 2025m12. Bins to the left of the red vertical line are pre-period bins, and bins to the right are post-ChatGPT bins. Each bin covers four months, while the endpoint bins cover two months each. The figure uses T-1 as the baseline period, which corresponds to July to October 2022. Vertical bars show 95% confidence intervals based on standard errors clustered at the occupation level. The estimates use the same sample, exposure measure, fixed effects, demographic controls, and weights as Table 2.

Table 3: Women/Men DID Source Diagnostics

	Women DID		Men DID	
	(1)	(2)	(3)	(4)
Usual weekly hours	-0.630***	--	-0.103	--
	(0.219)	--	(0.198)	--
Actual weekly hours	--	-1.339***	--	-0.641***
	--	(0.311)	--	(0.224)
Fixed effects	Occupation, province, month	Occupation, province, month	Occupation, province, month	Occupation, province, month
Observations	1,070,651	1,070,651	1,062,497	1,062,497

Note: This table reports DID estimates for two gender samples. Columns (1) and (2) represent the women sample, and columns (3) and (4) represent the men sample. Both samples are drawn from prime-age (age 25-54) LFS respondents in the paid employee labour force. The outcomes follow Table 2 and are usual weekly hours and actual weekly hours. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DID coefficient on AI exposure × post-ChatGPT. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. Each regression includes occupation, province, and month fixed effects. Each regression also includes demographic controls for age group, education, and immigrant status and uses LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Table A.1: AI Exposure by Occupation Group

Occupation group	Groups (n)	Weighted share (%)	AI Exposure (Anthropic) mean	AI Exposure (Anthropic) SD	AI Exposure (Eloundou) mean	AI Exposure (Eloundou) SD
Business, finance, and administration	5	18.4%	0.739	0.129	0.809	0.079
Health	5	8.6%	0.079	0.048	0.345	0.111
Management	4	7.6%	0.298	0.077	0.598	0.028
Natural and applied sciences	4	10.8%	0.569	0.387	0.792	0.202
Natural resources and agriculture	2	1.7%	0.033	0.030	0.090	0.105
Arts, culture, recreation, and sport	4	1.9%	0.389	0.220	0.618	0.194
Education, law, social, and government	8	14.0%	0.397	0.205	0.534	0.099
Manufacturing and utilities	3	4.6%	0.033	0.013	0.214	0.117
Sales and service	4	17.6%	0.272	0.150	0.364	0.166
Trades, transport, and equipment operators	4	14.7%	0.047	0.024	0.174	0.047

Note: This table reports AI exposure summaries by broad occupation group. The groups are the 10 broad categories in the National Occupational Classification reported in the LFS occupation coding. The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force observed in the pre-period from 2021m01 to 2022m10. Groups (n) is the number of detailed occupation groups in each broad occupation group. Weighted share is the broad occupation group's share of pre-period LFS observations after applying LFS final weights. Both AI exposure measures are assigned at the occupation level and normalized to [0,1]. Means and standard deviations are computed within each broad occupation group using LFS final weights.

Table A.2: Alternative Exposure

	(1)	(2)
	Usual weekly hours	Actual weekly hours
AI exposure × post-ChatGPT × female	-1.362***	-1.840***
	(0.333)	(0.537)
Occupation × month FE	Yes	Yes
Province × month FE	Yes	Yes
Female interacted FE	Yes	Yes
Demographic controls	Yes	Yes
Observations	2,133,148	2,133,148

Note: This table reports DDD estimates of how the post-ChatGPT change in the gap in weekly hours between women and men varies with AI exposure, using an alternative exposure measure. The gap is defined as women minus men. The outcomes are usual weekly hours in column (1) and actual weekly hours in column (2). The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2025m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 onward. AI exposure is measured at the occupation level using the Eloundou measure and is normalized to [0,1]. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status. All columns use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: FE Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Usual weekly hours				Actual weekly hours			
AI exposure × post-ChatGPT × female	-0.808***	-0.794***	-0.797***	-0.803***	-0.990**	-0.967**	-0.967**	-0.956**
	(0.292)	(0.290)	(0.292)	(0.293)	(0.406)	(0.414)	(0.416)	(0.418)
Occupation × month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation × female FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month × female FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province × month FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Province × female FE	No	No	Yes	Yes	No	No	Yes	Yes
Demographic controls	No	No	No	Yes	No	No	No	Yes
Observations	2,133,148	2,133,148	2,133,148	2,133,148	2,133,148	2,133,148	2,133,148	2,133,148

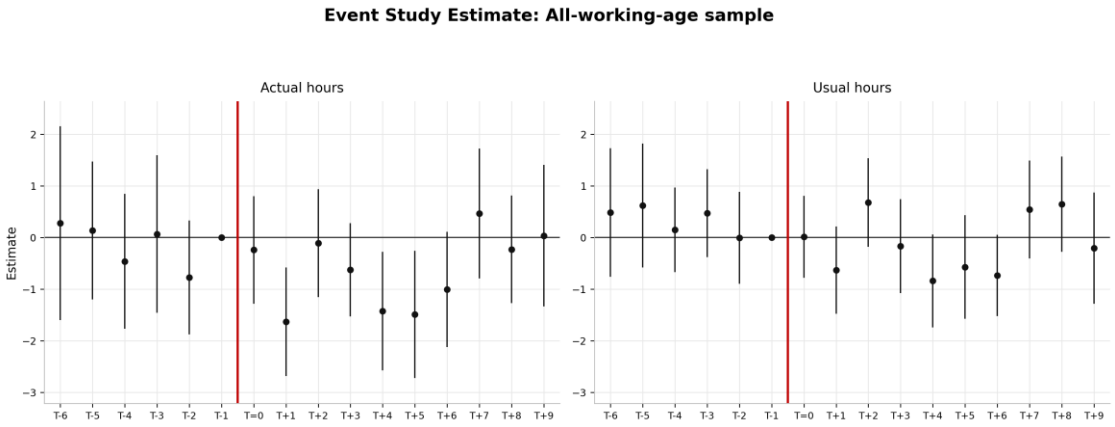
Note: This table reports DDD estimates of how the post-ChatGPT change in the gap in weekly hours between women and men varies with AI exposure, adding fixed effects and controls in steps. The gap is defined as women minus men. The outcomes are usual weekly hours in columns (1) to (4), and actual weekly hours in columns (5) to (8). The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2025m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 onward. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All columns include occupation-by-month, occupation-by-female, and month-by-female fixed effects, and each column estimates only the DDD coefficient. Columns (2) and (6) add province-by-month fixed effects. Columns (3) and (7) add province-by-female fixed effects. Columns (4) and (8) add demographic controls for age group, education, and immigrant status. All columns use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: All-Working-Age Sample Comparison

	(1)	(2)	(3)	(4)	(5)	(6)
	All-working-age		Prime-age		All-working-age excluding prime-age	
	Usual weekly hours	Actual weekly hours	Usual weekly hours	Actual weekly hours	Usual weekly hours	Actual weekly hours
AI exposure × post-ChatGPT × female	-0.385**	-0.494**	-0.803***	-0.956**	0.389	0.136
	(0.179)	(0.244)	(0.293)	(0.418)	(0.357)	(0.401)
Occupation × month FE	Yes	Yes	Yes	Yes	Yes	Yes
Province × month FE	Yes	Yes	Yes	Yes	Yes	Yes
Female interacted FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,164,118	3,164,118	2,133,148	2,133,148	1,030,970	1,030,970

Note: This table reports DDD estimates across three samples. Columns (1) and (2) represent the all-working-age (age 15-64) sample. Columns (3) and (4) represent the prime-age (age 25-54) sample and reproduce Table 2 columns (2) and (4). Columns (5) and (6) represent the all-working-age sample excluding the prime-age sample (age 15-24 and age 55-64). The outcomes are usual weekly hours in columns (1), (3), and (5), and actual weekly hours in columns (2), (4), and (6). All samples are drawn from LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2025m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 onward. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. The same specification is used in all six columns; only the sample changes across column pairs. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status and use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: All-Working-Age Event Study



Note: This figure plots event time estimates for the DDD coefficient in the all-working-age specification used in columns (1) and (2) of Table A.4. The left panel shows estimates for actual weekly hours, and the right panel shows estimates for usual weekly hours. The horizontal axis shows event time bins, and the vertical axis shows estimated coefficients in weekly hours. The event window runs from 2021m01 to 2025m12. Bins to the left of the red vertical line are pre-period bins, and bins to the right are post-ChatGPT bins. Each bin covers four months, while the endpoint bins cover two months each. The figure uses T-1 as the baseline period, which corresponds to July to October 2022. Vertical bars show 95% confidence intervals based on standard errors clustered at the occupation level. The estimates use the all-working-age (age 15-64) LFS respondents in the paid employee labour force, with the same exposure measure, fixed effects, demographic controls, and weights as columns (1) and (2) of Table A.4.

Table A.5: Joint Tests of Pre-Period Event Time Coefficients

Outcome	Joint pretrend p-value	Observations
Actual Weekly Hours	0.895	2,133,148
Usual Weekly Hours	0.641	2,133,148

Note: This table reports joint tests for the pre-period DDD event time estimates in the Figure 1 specification. The tests use the same event time bins as Figure 1. The event window runs from 2021m01 to 2025m12. Each bin covers four months, while the endpoint bins cover two months each. The baseline period is T-1, which corresponds to July to October 2022. The tests use the same sample, exposure measure, fixed effects, demographic controls, and weights as Table 2 and Figure 1. The reported p-values are from Wald tests. Standard errors are clustered at the occupation level.

Table A.6: Hours Scale Robustness

Outcome	Coef.	SE	N
Usual weekly hours	-0.803***	0.293	2,133,148
Actual weekly hours	-0.956**	0.418	2,133,148
log(1 + usual weekly hours)	-0.047***	0.017	2,133,148
log(1 + actual weekly hours)	-0.062*	0.031	2,133,148
IHS usual weekly hours	-0.053***	0.019	2,133,148
IHS actual weekly hours	-0.070*	0.036	2,133,148

Note: This table reports DDD estimates using alternative scales for weekly hours. The first two rows reproduce the level estimates from columns (2) and (4) of Table 2. Rows 3 and 4 use log(1 + hours). Rows 5 and 6 use the inverse hyperbolic sine transformation. For all rows, unemployed respondents in the relevant sample are assigned zero hours before transformation. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The sample, exposure measure, fixed effects, demographic controls, and weights are the same as in Table 2. The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All rows include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All rows add demographic controls for age group, education, and immigrant status and use LFS final weights. Standard errors are reported in the SE column and are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Placebo / Falsification

	(1)	(2)	(3)	(4)
	Fake event 2022m05		Fake event 2021m11	
	Usual weekly hours	Actual weekly hours	Usual weekly hours	Actual weekly hours
Fake event month	2022m05	2022m05	2021m11	2021m11
AI exposure × placebo post × female	-0.547	-0.304	-0.181	-0.132
	(0.417)	(0.699)	(0.459)	(0.568)
p-value	0.197	0.666	0.695	0.817
Occupation × month FE	Yes	Yes	Yes	Yes
Province × month FE	Yes	Yes	Yes	Yes
Female interacted FE	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Observations	423,510	423,510	703,217	703,217

Note: This table reports placebo DDD estimates using fake event dates. The outcomes are usual weekly hours in columns (1) and (3), and actual weekly hours in columns (2) and (4). The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the placebo DDD coefficient on AI exposure × placebo post × female. The gap is defined as women minus men. Columns (1) and (2) use a fake event month of 2022m05, with placebo post period 2022m05 to 2022m10. Columns (3) and (4) use a fake event month of 2021m11, with placebo post period 2021m11 to 2022m10. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status. All columns use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Sample Window Ending in 2024m12

	(1)	(2)
	Usual weekly hours	Actual weekly hours
AI exposure × post-ChatGPT × female	-0.856**	-1.188**
	(0.321)	(0.453)
Occupation × month FE	Yes	Yes
Province × month FE	Yes	Yes
Female interacted FE	Yes	Yes
Demographic controls	Yes	Yes
Observations	1,664,137	1,664,137

Note: This table reports DDD estimates of how the post-ChatGPT change in the gap in weekly hours between women and men varies with AI exposure, estimated on a shorter window. The gap is defined as women minus men. The outcomes are usual weekly hours in column (1) and actual weekly hours in column (2). The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2024m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 to 2024m12. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status. All columns use LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Cluster Influence and CV3J Inference

	(1)	(2)	(3)	(4)
	Usual hours		Actual hours	
	Anthropic	Eloundou	Anthropic	Eloundou
Panel A. Cluster influence				
Full sample estimate	-0.803	-1.362	-0.956	-1.840
Omit one estimate range	[-0.899, -0.603]	[-1.509, -1.125]	[-1.255, -0.683]	[-2.317, -1.533]
Same sign as full estimate	43/43	43/43	43/43	43/43
Panel B. CV1 and CV3J inference				
CV1 p-value	0.009	<0.001	0.027	0.001
CV3J p-value	0.011	<0.001	0.041	0.004
Observations	2,133,148	2,133,148	2,133,148	2,133,148
Occupation clusters	43	43	43	43

Note: This table reports cluster influence checks and CV3J inference following MacKinnon et al. (2023). Panel A omits one occupation cluster at a time. Panel B reports conventional CV1 p-values and CV3J p-values clustered at the occupation level. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. AI exposure is measured at the occupation level using the Anthropic measure in columns (1) and (3), and the Eloundou measure in columns (2) and (4). Both measures are normalized to [0,1]. The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status and use LFS final weights.

Table A.10: Main DDD Estimates with Demographic Composition Reweighting

	(1)	(2)	(3)	(4)
	Usual weekly hours	Usual weekly hours	Actual weekly hours	Actual weekly hours
AI exposure × post-ChatGPT × female	-0.803***	-0.846***	-0.956**	-1.002**
	(0.293)	(0.294)	(0.418)	(0.425)
Occupation × month FE	Yes	Yes	Yes	Yes
Province × month FE	Yes	Yes	Yes	Yes
Female interacted FE	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Demographic composition reweighting	No	Yes	No	Yes
Observations	2,133,148	2,133,148	2,133,148	2,133,148

Note: This table reports DDD estimates of how the post-ChatGPT change in the gap in weekly hours between women and men varies with AI exposure, with and without demographic composition reweighting. The gap is defined as women minus men. The outcomes are usual weekly hours in columns (1) and (2), and actual weekly hours in columns (3) and (4). Columns (1) and (3) reproduce columns (2) and (4) of Table 2. The sample is prime-age (age 25-54) LFS respondents in the paid employee labour force. For all hours outcomes, unemployed respondents in the relevant sample are assigned zero hours. The coefficient of interest is the DDD coefficient on AI exposure × post-ChatGPT × female. The estimation window is 2021m01 to 2025m12; the pre-period is 2021m01 to 2022m10, and the post-ChatGPT indicator equals one from 2022m11 onward. AI exposure is measured at the occupation level using the Anthropic measure and is normalized to [0,1]. All columns include the saturated DDD fixed effects, occupation-by-month, occupation-by-female, and month-by-female fixed effects, plus province-by-month and province-by-female fixed effects. All columns add demographic controls for age group, education, and immigrant status. Columns (1) and (3) use LFS final weights. The reweighting in columns (2) and (4) is calculated separately for women and men after pooling all occupation groups within each period. Post-period LFS final weights are adjusted so that each gender's joint distribution of age group, education, and immigrant status matches its pre-period distribution. Pre-period observations retain their LFS final weights. Standard errors in parentheses are clustered at the occupation level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.