Cognitive-Level Heterogeneity in the Employment Effects of Routine Intensity and AI Exposure: Evidence from the Diffusion of AI

Deokjae Jeong

Under review, *Korean Journal of Labor Economics*

Abstract

Using routine intensity and AI exposure, this paper analyzes how employment shifted across cognitive levels during the diffusion of AI. This study first construct the data by using a large language model to score O*NET task descriptions, obtaining granular routine and cognitive scores at the 6-digit SOC occupation level. The paper then analyze U.S. employment with a two-stage Difference-in-Differences design that treats routine intensity as a continuous treatment and uses the cognitive score to split the sample. The sign of the post-treatment coefficient splits at the cognitive threshold. The negative coefficient in the low-cognitive group is not a new, AI-driven phenomenon but is explained by the continued pattern of Routine-Biased Technological Change, whereas the positive coefficient in the high-cognitive group suggests that, contrary to what Routine-Biased Technological Change predicts, routine occupations are gaining ground within the high-cognitive group. When the same design instead uses AI exposure as the treatment, the post-treatment coefficient is not statistically significant for the low-cognitive group, whereas for the high-cognitive group it is estimated to be negative —opposite in sign to the case using routine intensity as the treatment. Since these two measures, routine intensity and AI exposure, are uncorrelated, they are interpreted as offering different perspectives on the labor market. This suggests that the routine measure on which the Routine-Biased Technological Change literature has relied does not represent the mechanism through which AI affects employment.

Introduction

The rapid diffusion of generative artificial intelligence, following the release of conversational models in late 2022, has thrust the question of which jobs the technology will affect to the forefront of labor economics. Generative AI is closer to a capacity for human-like reasoning than to a technology that merely substitutes for routine work requiring nothing more than rule-following; if so, it may reach more deeply into the low-routine, non-standard tasks that predominate in high-cognitive occupations. By contrast, there is little disagreement that low-cognitive occupations —those relying on the hands and the body— are difficult for generative AI to replace and should therefore see little change in employment. This study examines these expectations empirically using the U.S. Current Population Survey, but frames its question more broadly rather than asking simply how much generative AI has reduced employment. The question this study seeks to answer is how the relationship between an occupation's routine intensity and its level of employment shifted across occupations of differing cognitive levels over the period around 2023 —when generative AI began to spread in earnest— and what explains that shift. It further asks how the answer changes when AI exposure, rather than routine intensity, is used to pose the same question.[1] Routine intensity denotes standardized, automatable tasks as it approaches 1 (routine), and non-standard tasks that call for handling the unexpected as it approaches 0 (non-routine). Cognitive intensity denotes tasks centered on mental processing as it approaches 1 (cognitive), and physical tasks performed with the hands and the body as it approaches 0 (manual).

To answer the question it raises, this study proceeds in three steps. First, it measures the routine and cognitive content of occupations in a new way. Prior research on Routine-Biased Technological Change (hereafter RBTC) has been aware of a measurement problem: conclusions about employment effects hinge on how an occupation's Routine Task Intensity (hereafter RTI) is measured (Walo 2023; Haslberger 2022). In response, this study has a large language model (LLM) directly interpret O*NET's[2] occupation-level task descriptions and score them, on a 0-to-1 scale, into routine and cognitive scores. To the author's knowledge, this is the first attempt to apply an LLM to the measurement of routine task intensity. Measuring routine and cognitive scores in this way permits far more finely disaggregated occupation-level measurement than the approaches used in the existing literature. Second, treating the resulting routine intensity as a time-invariant, continuous treatment, it estimates a two-stage Difference-in-Differences. Specifically, because the long-run employment trend differs across occupation groups, the first stage strips out each occupation group's long-run trend to obtain detrended employment. Using this as the dependent variable, the second stage then estimates period-specific dose-response coefficients within an event-study framework. When a post-treatment coefficient is estimated below zero, this paper refers to it as a negative (−) coefficient, and when above zero, as a positive (+) coefficient. To obtain accurate standard errors throughout the procedure, an occupation-group cluster bootstrap is used. Third, splitting the full sample into two groups at a cognitive score of 0.5, it compares how the post-treatment coefficients on routine intensity diverge between the low-cognitive group (primarily physical tasks) and the high-cognitive group (primarily mental tasks). On an equal footing with routine intensity, it also uses the AI exposure index as the treatment in the same two-stage difference-in-differences, comparing how the two measurement axes

---

[1] Recent research linking generative AI and labor has concentrated on measuring which occupations are exposed to AI (Eloundou et al. 2024; Webb 2019; Felten, Raj, and Seamans 2021). An exposure index by itself indicates only that an occupation has considerable potential to be affected; it does not encode whether that potential will play out as an increase or a decrease in employment. When this study uses the exposure index as the treatment intensity in a difference-in-differences analysis, it shows that—at least when the sample is split by cognitive level—employment declines more sharply in occupations with greater AI exposure within the high-cognitive group. That said, even as generative AI is being adopted rapidly (Bick, Blandin, and Deming 2025; Humlum and Vestergaard 2024), assessments continue to hold that the labor-market effects during the early diffusion phase are generally small or only weakly identified (del Rio-Chanona et al. 2025); and when AI-unrelated shocks overlap in the same period, attributing outcomes to AI as a single cause may be hard to defend.

[2] O*NET (Occupational Information Network) is an occupational-information database sponsored by the Employment and Training Administration (ETA) within the U.S. Department of Labor. It provides standardized information—on knowledge, skills, and abilities; tasks; and work context—for more than 900 occupations.

illuminate employment change differently.

The study's core findings can be stated at the outset. The sign of the post-treatment coefficient on routine intensity splits at the cognitive threshold. In the low-cognitive group, where mental tasks are few, more routine occupations experienced a relatively steeper decline in detrended employment after treatment, yielding a negative (−) post-treatment coefficient. In the high-cognitive group, where mental tasks are plentiful, more routine occupations instead saw their employment improve in relative terms, yielding a positive (+) coefficient. Moreover, the two groups' post-treatment coefficients are not merely opposite in direction but also lopsided in magnitude: the negative (−) coefficient for the low-cognitive group is two to three times as large as the positive (+) coefficient for the high-cognitive group. It bears emphasizing, however, that negative and positive here refer not to absolute gains or losses in jobs but to relative employment magnitudes once each occupation group's long-run employment trend has been removed. That is, even a positive (+) post-treatment coefficient need not signal that employment rose; it may signal only that employment did not fall by as much as the prior trend would have implied.

This study does not treat this sign reversal as the effect of generative AI alone; rather, it disentangles where each of the two groups' divergent post-treatment coefficients originates. First, the negative (−) post-treatment coefficient observed in the low-cognitive group is nothing new. It is explained fairly naturally as a continuation of the long-standing pattern of RBTC, in which routine, physical tasks have been pushed aside by machines and software. Ever since the early work that viewed labor through the lens of RBTC (Autor, Levy, and Murnane 2003),[3] rule-governed routine tasks have been understood as readily transferable to machines, and it has been widely documented that this substitution has eroded the routine occupations of middle-skill workers and polarized the labor market (Autor and Dorn 2013). Studies that examine robots and physical automation directly likewise confirm that this displacement effect[4] is concentrated in routine, physical jobs (Acemoglu and Restrepo 2020; Dauth et al. 2017).[5] The relative contraction of low-cognitive, high-routine occupations is thus more naturally read as the reappearance —once again along the routineness channel— of forces that were already at work before generative AI.

Conversely, it is uncertain whether the positive (+) post-treatment coefficient observed in the high-cognitive group is attributable to AI. Routineness gauges only how regular or irregular a task is; it does not map one-to-one onto whether generative AI can substitute for that task. Indeed, indices that quantify how exposed an occupation is to advances in AI are reported to take higher values for non-routine mental tasks (Eloundou et al. 2024; Webb 2019; Felten, Raj, and Seamans 2021). This study's own analysis finds that routine intensity and AI exposure intensity are

---

[3] More precisely, they described computer automation as "substituting for" high-cognitive routine tasks and low-cognitive routine tasks, and as "complementing" high-cognitive non-routine tasks and low-cognitive non-routine tasks.

[4] This is the term used by Acemoglu and Restrepo (2020), a concept distinct from the substitution effect and the reinstatement effect.

[5] More precisely, Acemoglu and Restrepo and Dauth et al. (2020) characterized the negative employment effects of automation as concentrated mainly in manufacturing, low-cognitive routine occupations, blue-collar jobs, and assembly-line occupations.

negatively correlated across the full sample, but exhibit no correlation whatsoever once attention is restricted to the high-cognitive group. It therefore remains an open question why, within the high-cognitive group, employment fell in occupations with greater AI exposure while it rose in occupations with greater routineness. Precisely because the two intensities are uncorrelated within the high-cognitive group to begin with, it is risky to declare that the routine result is due solely to AI.

The finding that more routine occupations gained employment within the high-cognitive group suggests that, since AI began to be adopted in 2023, a phenomenon at odds with the mechanism RBTC has long described has been under way. Under RBTC, it is generally the case that —regardless of a worker's cognitive level— more routine occupations are displaced by automation technologies and suffer a decline in their employment share. Yet since 2023, the high-cognitive group instead exhibits a heterogeneous change that departs from this established theoretical prediction, with more routine occupations gaining ground in employment. Whether this reversal in the employment trajectory stems entirely from the arrival of generative AI cannot be asserted causally, but the timing of the structural change coincides exactly with the AI diffusion period. This alignment strongly supports the possibility that AI is acting on high-cognitive routine tasks as a powerful complementary productivity shock —one different in kind from earlier computer capital.

When the AI exposure index is substituted for routine intensity as the treatment variable in the same difference-in-differences design, a markedly different picture emerges. In the high-cognitive group, the post-treatment coefficient is estimated to be negative (−): occupations with greater AI exposure experienced more adverse employment effects. This is the exact opposite sign to the positive (+) obtained when routine intensity served as the treatment variable. But because the two measures are effectively independent under the high-cognitive condition, the opposing negative (−) and positive (+) results cannot be attributed to the routine index being a mirror image of the AI exposure index. Rather, the routine index and the AI exposure index illuminate different aspects of the labor market. In the low-cognitive group, meanwhile, the post-treatment coefficient is not significantly different from zero —there is no effect at all. This is consistent with the commonsense expectation that AI would have no effect on occupation groups that make heavy use of physical labor.

The remainder of the paper is organized as follows. The next section, "Data Construction," describes the dependent variable —employment— together with the LLM-scored routine and cognitive scores and the second treatment axis, the AI exposure index. "Difference-in-Differences Analysis" builds a two-stage, continuous-treatment difference-in-differences model that removes the occupation-group-specific long-run employment trends, and then presents the results in three parts: first, under routine intensity as the treatment, how the sign of the post-treatment coefficient splits by cognitive level (the sign reversal); next, placing the AI exposure index as the treatment in the same design, a negative (−) post-treatment coefficient for the high-cognitive group and no significance for the low-cognitive group. "Robustness Checks" uses the wild cluster bootstrap to show that the core results along both axes —routine intensity and AI exposure— hold up even when the inference method is changed. Finally, the "Conclusion" draws together the findings and their implications.