

Makridis, Christos

Working Paper

The Labor Market Effect of Generative Artificial Intelligence on Artists

CESifo Working Paper, No. 12368

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Makridis, Christos (2025) : The Labor Market Effect of Generative Artificial Intelligence on Artists, CESifo Working Paper, No. 12368, Munich Society for the Promotion of Economic Research - CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/336068>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

CES ifo

12368
2025

December 2025

Working Papers

The Labor Market Effect of Generative Artificial Intelligence on Artists

Christos Makridis

CES ifo

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

The Labor Market Effect of Generative Artificial Intelligence on Artists

Christos A. Makridis *

December 26, 2025

Abstract

Technological change has repeatedly disrupted creative labor markets, raising concerns about whether new tools substitute for artists or shift the organization of creative work. This paper studies how occupational exposure to generative AI (genAI) maps into employment and earnings outcomes for U.S. artists following the unanticipated release of ChatGPT. I combine an occupation-level LLM task exposure index with establishment-based occupational outcomes from the Occupational Employment and Wage Statistics (OEWS) and individual microdata from the American Community Survey (ACS), estimating event-study specifications that compare more versus less exposed artistic occupations from 2017 to 2024. Across datasets, I find little evidence of short-run earnings declines associated with LLM exposure through 2023, with point estimates near zero and in some specifications modestly positive. Evidence on employment is more mixed, with weaker employment growth in 2023 for more exposed artistic occupations in some specifications. To investigate mechanisms, I use the Gallup Workplace Panel from 2023 to 2025 to measure AI use directly and relate changes in AI use to job satisfaction and burnout. Within-person estimates show limited average well-being effects of adoption, but suggest heterogeneous responses for artists and a use pattern concentrated in ideation and creative-support tasks. The results are consistent with early task reallocation than immediate labor-market harm, while leaving open the possibility of medium-run adjustment as adoption deepens and complementary investments accumulate.

Keywords: artificial intelligence; arts; culture; creative economy; large language models; generative AI; skills dynamics; workforce.

JEL Codes: J23; J24; L82; O33

*Christos: Arizona State University, Gallup, and Stanford University, cmakridi@asu.edu.

1 Introduction

Technological shocks have repeatedly disrupted creative work. When photography emerged in the nineteenth century, many observers predicted that painting would become obsolete, and portrait commissions did fall as photographs produced realism more cheaply.¹ Yet painting adapted, with artists shifting toward styles that emphasized interpretation rather than replication. A parallel dynamic appeared in recorded music after MP3s and file sharing: revenues fell sharply, but the industry reorganized around new distribution and performance models (Waldfoegel, 2017). These episodes motivate the central question of this paper: is generative AI different for the creative economy, and what are its causal effects on employment and earnings in creative occupations?

Large language models and other generative AI systems have intensified debate about their effects on creative labor markets. Creative professionals raise two main concerns. First, many models are trained on copyrighted text, images, and music, often without permission, creating legal and compensation disputes and the risk of infringement in generated outputs.² Second, there is concern about displacement as these tools automate or partially automate tasks in illustration, writing, translation, and composition.³ While earlier work suggested that creative occupations might be relatively resistant to automation because they rely on open ended and original work, more recent evidence points in the opposite direction. Eloundou et al. (2024) estimate that exposure to LLM capabilities is widespread across the workforce, and it is often higher in knowledge

¹Contemporary reactions to the daguerreotype frequently framed photography as an existential threat to painting, with critics predicting the obsolescence or death of traditional art forms (McCouat, 2015).

²For instance, a high-profile lawsuit by Getty Images alleges that an AI company “misus[ed] more than 12 million [copyrighted] photos to train its...image-generation system,” in violation of IP law (Brittain, 2023). Artist groups filed class-action suits claiming that generative AI tools were built on uncompensated use of their work.

³Frey and Osborne (2017) estimated a high automation risk for routine jobs but a much lower risk for artistic and creative ones, positing creativity as a “bottleneck” to computerization. Indeed, Bakhshi et al. (2015) projected that about 21% of U.S. employment falls in highly creative occupations largely safe from automation.

intensive and higher wage occupations, including roles in design, media, and the arts.

However, the results may also be positive. In particular, [Epstein and Hertzmann \(2023\)](#) write that “generative AI is not necessarily the harbinger of art’s demise, but rather is a new medium with its own distinct affordances” by reducing the barriers to entry and leading to an expanded set of new tasks, products, and services. And yet, little empirical evidence exists. [Zhou and Lee \(2024\)](#) provide some optimism: studying over 4 million artworks from more than 50,000 unique users, they find that text-to-image AI raises human creative productivity by 25% and increases the value as measured by the likelihood of receiving a favorite per view by 50%. They point out, however, that there could be some consequences: average novelty among the content declines.

The primary contribution of this paper is to estimate how exposure to large language models maps into labor market outcomes in the creative economy, with a focus on professional artists. I exploit plausibly exogenous variation from the unanticipated introduction of ChatGPT in 2022 and combine it with cross occupational heterogeneity in LLM task exposure to study changes in employment and earnings. Using both state level occupational data from the Occupational Employment and Wage Statistics (OEWS) and individual microdata from the American Community Survey (ACS) from 2017-2024, I examine whether more exposed artistic occupations experienced differential shifts in unemployment, employment growth, and earnings after 2022, controlling for demographics and fixed effects. Across specifications, I find limited evidence of earnings losses associated with LLM exposure in the short run and, if anything, positive effects. An additional standard deviation increase in exposure is largely uncorrelated with earnings through 2023, with point estimates that are modestly positive but not statistically distinguishable from zero. By contrast, there is some evidence of weaker employment growth in 2023 among more exposed artistic occupations. These patterns are robust to a triple difference design that compares artists to a

broader arts, entertainment, and sports group, combined with coarsened exact matching on age and education to improve comparability (Iacus et al., 2011, 2012).

To shed light on the mechanisms behind these effects, I draw on the Gallup Workplace Panel, a nationally representative longitudinal survey from 2023 to 2025 that measures both the frequency and purpose of AI use. Although the artist subsample is small and within person variation is limited, the evidence suggests that AI adoption in the arts operates through task reorganization rather than broad improvements in well being. Frequent AI use is not associated with systematic gains in job satisfaction once individual fixed effects are introduced, and its relationship with well being differs by how artists are defined: intensive use is linked to lower job satisfaction for industry defined artists, but to lower burnout for occupation defined artists. Reported use cases further clarify this pattern. Relative to the broader workforce, artists are substantially more likely to use AI for idea generation and creative support tasks, and less likely to use it for customer facing or operational functions. These findings are consistent with a cautiously optimistic interpretation in which generative AI complements upstream creative work and workflow coordination, while also reshaping the content and experience of artistic labor in heterogeneous ways.

A further reason for caution is that the labor market effects identified here likely reflect an early phase of diffusion rather than a settled long run equilibrium. A recurring theme in the literature is that measured employment and wage responses can lag adoption because the main productivity and displacement margins require complementary investments in intangible capital, organizational redesign, and new contracting and governance arrangements (Makridis, 2025d), producing nonlinear adjustment paths rather than immediate level shifts (Brynjolfsson et al., 2021). In the context of LLMs, this implies that short run null effects on earnings are compatible with meaningful medium run reallocation across tasks, occupations, and firms as adoption deepens and

complementary changes accumulate. Recent evidence also suggests that the aggregate employment response to AI to date has been modest and uneven, with detectable effects concentrated in specific exposure intensive segments rather than broad based disemployment (Humlum and Vestergaard, 2025), consistent with an augmentation channel dominating in early stages while substitution pressures remain localized on sub-groups (Brynjolfsson et al., 2025). Nonetheless, there is some evidence from administrative data that employment and wages can benefit as adoption becomes more widespread and integrated into production, reinforcing the possibility that impacts intensify over time even if early average effects are small (Johnston and Makridis, 2025). The literature is starting to move away from simply estimating positive or negative marginal effects of genAI exposure towards answering questions over who benefits and how.

This paper contributes most closely to an emerging literature on exposure to AI and the effects of technological change on work (Acemoglu and Restrepo, 2019). While automation-based technologies have historically been labor-displacing (Acemoglu and Restrepo, 2020), there has been a lot of debate about whether LLMs have the potential to augment human activity. Even if LLMs lead to the displacement of some tasks, and more likely a transformation of work within jobs (Eloundou et al., 2024), they could still have positive effects. For example, Zhou and Lee (2024) find that text-to-image AI raises human creative productivity, boosting the number of artworks produced by about 25–50% within a few months of adoption. They also found that the AI-assisted works attracted more audience appreciation on average, suggesting a potential complementary effect where human creators leverage AI to amplify their reach. However, others argue that AI could allow a few superstar creators (or firms) to vastly increase output, it might reduce overall demand for other artists, concentrating earnings among fewer winners (Acemoglu, 2021). In sum, the net effect of LLMs on creative work—whether a “great boon” in which human

artists become more productive, or a “big blow” that deskills and displaces creative labor—remains an open question. These concerns echo historical episodes in which new technology upset creative fields, initially threatening jobs and IP norms before eventually spurring adaptation.

This paper also contributes to a larger literature at the intersection of labor and cultural economics. Early empirical research debated whether artists truly earn less than comparable workers. Using U.S. Census data for 1980, [Filer \(1986\)](#) found that after controlling for education and other characteristics, the average artist earned roughly 10% less than a similar non-artist and that over a lifetime this earnings difference nearly vanished. Filer’s finding suggested that the “starving artist” stereotype might be overstated and that lifetime earnings of artists were almost on par with other workers, partly because those who remain in artistic careers tend to experience faster earnings growth (with less successful artists exiting the field). However, subsequent studies consistently documented an earnings penalty for artists. For example, [Alper and Wassall \(2006\)](#) argued that artists do suffer a pay disadvantage (though not usually outright poverty) relative to their qualifications. In a comprehensive analysis spanning 1950–2000, they estimated that the earnings shortfall for artists varied widely across decades – from as low as about 6% to as high as 51% less than comparable professionals – depending on the time period. More recent evidence suggests the earnings gap may have worsened in the 21st century. [Makridis \(2025c\)](#) analyzes American Community Survey data from 2006–2021 and finds a decline in artists’ relative earnings from near parity in the mid-2000s to roughly a 15% shortfall by the 2010s. After adjusting for demographic differences, the earnings disadvantage of artists grew even sharper at a 30% deficit. [Makridis \(2025c\)](#) also finds that individuals with fine arts degrees face an earnings penalty and lower employment rates, even when they work in artistic jobs, but it is partially offset through exposure to business education ([Makridis and Kuuskoski, 2024](#)). This paper adds to the literature

by providing some optimistic evidence that LLMs could have positive effects for artists.

2 Data and Measurement

2.1 State Occupational Employment Data

The Occupational Employment and Wage Statistics (OEWS) is an annual survey program run by the U.S. Bureau of Labor Statistics that provides detailed estimates of employment and earnings by occupation. The OEWS is a large survey of business establishments (employers). It covers wage and salary workers in non-farm industries, across approximately 800 SOCs. For each occupation, OEWS reports the estimated number of people employed and the mean and percentile earnings (hourly and annual). The OEWS data are highly reliable for occupational employment statistics. The sample is large (about 1.1 million establishments over three years) and designed to produce accurate estimates at national, state, and metropolitan levels, or by industry. A key limitation of OEWS, however, is that it excludes several categories of workers important to the creative economy, namely the self-employed and owners of unincorporated firms. Many artists, writers, and performers work as freelancers or independent contractors; these individuals do not appear in OEWS because no establishment payroll survey covers them. Thus, OEWS may undercount total creative employment and miss earnings from freelance work.

2.2 Classifying Artists

I define artists based following the Standard Occupational Classification (SOC) system, namely:

(a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b)

actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209).⁴ This definition of artists is more narrow than other approaches on the arts and cultural sectors, largely because the lives of professional artists are systematically different than those of cultural sector workers who may have a salaried position. When turning to the Gallup Workplace Panel data (see Section 2.5), I will deviate from this classification due to data constraints.

2.3 Individual Earnings Data

My primary data comes from the American Community Survey (ACS) accessed through Integrated Public Use Microdata (Ruggles et al., 2024).⁵ The ACS is an annual (rolling) survey conducted by the U.S. Census Bureau that samples over 3 million individuals each year, collecting detailed information on demographics, socioeconomic status, and housing. For our purposes, the ACS provides microdata on workers in creative occupations, including their employment status, income, and personal characteristics. The ACS also asks respondents about their current occupation (for employed individuals) and class of worker (private company, government, self-employed, etc.), as well as their income over the past 12 months from various sources. The ACS offers a rich picture of creative workers' economic outcomes. Because it includes self-employed and freelance workers, I can observe segments of the creative labor force that would otherwise be missed by establishment data, like the OEWS, offering a complementary glimpse. This is particularly important in fields

⁴Artists are defined as those in SOC codes 27-1010, 27-2010, 27-2030, 27-2040, and 27-2090— this covers fine artists, actors/directors, choreographers/dancers, musicians/singers, and other entertainers. Notably, I exclude designers (27-1020) and athletes (27-2020) from the artist definition to focus on creative art occupations.

⁵Unfortunately, no other data has the sample size to speak about this sub-group. The American Time Use Survey (ATUS), for example, only contains 63 observations for artists from 2019-2023. The monthly Current Population Survey (CPS) similarly only has 1,063 respondents. Both, however, have a larger barrier: artists are grouped together (code 2600), which does not give any heterogeneity across the different groups

like visual arts or music, where freelance gigs are common. The ACS also lets me break down outcomes by demographics (gender, race, education, age). I restrict the sample to those who are employed and earning over \$15,000 and 300 total hours worked per year.⁶

The micro-data allows me to examine heterogeneity in treatment effects by, for instance, age and education.⁷ Moreover, ACS is useful for capturing earnings in a broad sense – including secondary jobs or gig income that might not appear on a single employer’s payroll. For instance, if a graphic designer starts supplementing their income through online freelance platforms (potentially competing with AI tools), ACS income data would reflect that, whereas OEWS might not. Nevertheless, the ACS has sampling error and survey response issues. In addition, the occupational coding is based on respondents’ self-reported job titles and duties, which might be inconsistently reported or misclassified. For example, someone doing freelance multimedia art might just report “artist” or might report “self-employed consultant,” depending on interpretation.

2.4 Large Language Model (LLM) Exposure Data

A key component of my empirical strategy relies on measures of occupational exposure to large language models (LLMs). To quantify this exposure, I adopt the methodology introduced by [Eloundou et al. \(2024\)](#), who construct a systematic exposure index that captures the susceptibility of different occupations to task automation or augmentation by LLMs. Specifically, their paper introduces a measure (denoted by the parameter β) which represents the share of an occupation’s tasks potentially affected by LLMs. Following a similar strategy as the suitability of machine learning (SML) index from [Brynjolfsson et al. \(2018\)](#), [Eloundou et al. \(2024\)](#) combined detailed

⁶The results become weaker if the cutoff is reduced—that is, the genAI effects look slightly more negative, but still not statistically different from zero, if the cutoff is, say, \$10,000/year.

⁷See Table A.1 in the Online Appendix with summary statistics separately for artists and non-artists.

occupational task descriptions from the Occupational Information Network (O*NET) database with expert human evaluations. Specifically, experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task. This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score.⁸

The parameter β can be interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. Each occupation's β thus reflects the fraction of tasks within that occupation that LLMs can either fully automate or meaningfully assist, making it a robust indicator of occupational vulnerability or adaptability to generative AI. Higher values of β indicate occupations where AI can readily substitute or complement human effort—such as text-based content creation or preliminary design work—while lower values imply occupations where tasks remain uniquely human-centric, like live performance or nuanced interpersonal interactions.

2.5 The Gallup Workplace Panel

To provide robustness and a more direct way of measuring AI adoption, I draw on the Gallup Workplace Panel, a probability-based longitudinal survey conducted by Gallup. It recruits a nationally representative sample of adults aged 18 and above using random sampling methods (address-based and random-digit dialing recruitment), ensuring coverage of the U.S. population. Starting in 2023, Gallup began surveying respondents annually about AI utilization and, in Q2:2025, Gallup began asking AI utilization questions quarterly. In particular, respondents are given a definition of AI:

⁸See Section A.1 of the Online Appendix for a detailed synthesis of their contribution. Table A.2 in the Online Appendix also provides a ranking of LLM exposure for each of the 11 artist occupations, ranging from music directors & composers at the top in terms of exposure to dancers and actors at the bottom.

“Artificial intelligence is when a computer does or enhances a task that is normally performed by a person.” Then, they are asked “How often do you use artificial intelligence in your role,” which include: (a) daily, (b) a few times a week, (c) a few times a month, (d) few times a year, (e) once a year, (f) less often than once a year, and (g) never. I consolidate these answers into three groups: often (daily and a few times a week), sometimes (a few times a month, a few times a year), and never (once a year, less often than once a year, and never).

I focus on two other outcomes: burnout and job satisfaction. The former is measured as follows: “Please indicate how often each of the following statements is true of your primary job (i.e., the job where you spend the most time working). I feel burned out at work: (a) Always, (b) Very often, (c) Sometimes, (d) Rarely, (e) Never.” The latter is measured (ranked on a 1 to 5 scale) as follows: “How satisfied are you with your place of employment as a place to work?”

My analytic file restricts to employed respondents (full- and part-time) and uses Gallup’s sampling weights. The pooled sample spans 2023-2025 and contains 70,877 person-wave observations. Because the 2025 sample includes repeated observations for some respondents, I report both the total number of person-wave observations and the number of unique respondents by year in Table A.4. To study artists within this workforce sample, I rely on two complementary tagging approaches: an industry-based definition (`artind`, Arts/Design/Entertainment/Sports/Media) and an occupation-based definition (`artocc`, SOC = 27, Arts, Design, Entertainment, Sports, and Media Occupations). These definitions provide parallel mappings of “artists” in the Gallup panel while preserving coverage across the broader labor market.

Table A.3 reports weighted summary statistics for the pooled sample and key subgroups. Relative to the pooled workforce (AI use: 0.19 often, 0.18 sometimes, 0.60 never; shares may not sum to one due to rounding), artist adoption differs by tagging approach: the occupation-defined

artists exhibit higher AI utilization (0.24 often and 0.27 sometimes, with 0.48 never), whereas the industry-defined artists look closer to the pooled mean (0.17 often, 0.20 sometimes, 0.61 never). Artists work fewer hours per week (34.9–35.6 versus 38.8 pooled) and are more likely to be part-time (0.34 for Arts(ind) and 0.26 for Arts(occ) versus 0.19 pooled), especially under the industry definition. Both artist groups are also more educated than the pooled workforce (college-or-more: 0.54 for Arts(ind) and 0.60 for Arts(occ) versus 0.43 pooled). Finally, the business/finance occupational benchmark displays markedly higher frequent AI use (0.42 often) and a substantially lower “never” share (0.32) than either artist category. Table A.4 shows that the pooled dataset contains 70,877 person-wave observations from 34,544 unique respondents, with a large increase in 2025 person-wave observations (51,589) reflecting repeated within-year measurement (28,922 unique respondents). Using the industry-based definition, the pooled sample includes 1,909 artist observations (1,464 unique individuals), while the occupation-based definition yields 1,797 artist observations (1,317 unique individuals), with most artist-tagged observations occurring in 2025.

3 Empirical Strategy

To understand the effect of generative AI on artist labor market outcomes, I exploit the introduction of ChatGPT in 2022 and the heterogeneous exposure of varying occupations:

$$y_{iot} = \delta^t \left(\sum_{t=2017}^{2024} EXPOSE_o \times Year_t \right) + \zeta X_{it} + \phi_o + \lambda_t + \varepsilon_{iot} \quad (1)$$

where y_{iot} denotes the outcome variable of interest for individual (or state) i , $EXPOSE_o$ denotes an index of occupational exposure to LLMs from Eloundou et al. (2024), $Year_t$ denotes year fixed effects (normalized to 2020), X denotes a vector of demographic covariates (when using

the Census Bureau data), and ϕ and λ denote fixed effects on occupation and year. The outcomes of interest are state \times occupation employment and wages from the Occupational Employment and Wage Statistics (OEWS) and individual annual earnings and hours worked from the American Community Survey (ACS). Standard errors are clustered at the occupational-level.

The identifying assumption in Equation 1 is a parallel trends style restriction: absent the diffusion of LLM-based tools, outcomes in more exposed and less exposed occupations would have evolved similarly, conditional on occupation and year fixed effects (and the individual controls in the ACS specifications). Equivalently, any unobserved shocks that affect employment, earnings, or hours may vary over time, but they must not be systematically correlated with the pre-measured exposure index in a way that changes differentially after 2022.

This assumption is considerably weaker than what would be required in a cross-sectional regression of outcomes on exposure. A purely cross sectional comparison would be vulnerable to non random sorting, for example if workers with higher latent productivity, different risk preferences, or stronger outside options are disproportionately represented in occupations with certain task content and therefore different measured exposure. The empirical strategy here instead uses the unanticipated timing of ChatGPT’s release in late 2022 as a common shock and asks whether outcomes evolve differentially over time as a function of pre-existing exposure. In this setting, the key threat is not static selection into occupations per se, but selection or reallocation that is correlated with exposure and occurs around the launch in a way that would have happened even absent generative AI. I assess the plausibility of this assumption in two ways. First, the event study specification allows visual and statistical inspection of pre-period exposure gradients, which are small and statistically indistinguishable from zero, consistent with no differential pre-trends. Second, to address the concern that exposure may proxy for time varying industry demand shocks,

I estimate versions of Equation 1 that include industry by year fixed effects in the ACS specifications. The results are qualitatively unchanged, although the standard errors are higher, suggesting that the main patterns are not driven by broad industry level shocks that differentially affect more exposed occupations over time—of particular concern during the pandemic.

As additional robustness, I also consider a triple-difference estimator that compares artists to non-artists, together with coarsened exact matching (CEM) weights that match individuals based on demographics (e.g., age and education) (Iacus et al., 2011, 2012):

$$y_{iot} = \xi^t \left(\sum_{t=2017}^{2024} EXPOSE_o \times ARTIST_i \times Year_t \right) + \zeta X_{it} + \phi_o + \lambda_t + \varepsilon_{iot} \quad (2)$$

Before turning to the main results, it is useful to clarify what the empirical estimates capture. The main specifications relate changes in outcomes after 2022 to pre measured occupational exposure to generative AI, with fixed effects absorbing time invariant differences across states, occupations, and individuals as applicable. A causal interpretation therefore requires that, absent generative AI, outcomes in higher and lower exposure occupations would have evolved similarly. The event study evidence supports this assumption in the pre period, but it cannot fully rule out contemporaneous shocks that are correlated with exposure. With that caveat, the central empirical pattern is that more exposed creative occupations experience a clearer increase in annual hours, alongside more modest changes in annual earnings, in the post 2022 period.

4 Main Results

Figure 1 reports event study estimates from the OEWS linking occupational exposure to generative AI to outcomes for artistic occupations, normalized to 2020. The estimates do not show

a sharp shift in the exposure gradient after 2020. Instead, coefficients are generally small and imprecise, with wide confidence intervals across outcomes. In Panels A and B, which examine levels of log earnings and log employment, pre-2020 coefficients do not show differential pre-trends and are statistically indistinguishable from zero, consistent with the identifying assumption that outcomes did not exhibit differential pre trends with respect to the exposure measure. After 2020, point estimates for earnings drift modestly upward and employment estimates are slightly positive on average, but the confidence intervals overlap zero throughout, indicating noise. If anything, however, there is a positive association between genAI exposure and labor market outcomes.

Panels C and D examine year to year changes as additional outcomes. Earnings growth is modestly positive in several post 2020 years, especially 2022 and 2023, but remains imprecisely estimated. Employment growth is more often negative, including in several pre 2020 years, which suggests that the level relationship between exposure and employment growth may differ across occupations even absent a generative AI shock. Interpreting the event study coefficients as changes relative to 2020, the OEWS evidence does not support large, contemporaneous declines in earnings or employment among more exposed artistic occupations in the short run.

Figure A.2 in the Online Appendix replicates the OEWS analysis using the human coded exposure measure and yields broadly similar, more muted estimates. Figure A.3 reports a triple interaction specification, plotting coefficients on artist by year by exposure terms, which isolates how the exposure gradient evolves for artists relative to non-artists over time. These estimates suggest somewhat stronger post 2020 earnings patterns and weaker employment growth, though uncertainty remains substantial. Results are similar when using state by occupation fixed effects.

A limitation of the OEWS is that it is establishment based and does not cover many freelance work arrangements that are common among artists. It also lacks demographic controls, so

compositional changes correlated with exposure may confound interpretation. To address these limitations, Figure 2 reports individual level ACS results. Panel A shows a modest increase in log earnings in 2023 for individuals in more exposed artistic occupations, which attenuates in 2024. The pre 2020 coefficients remain near zero. Panel B shows a positive association between exposure and log total hours beginning in 2022 and persisting through 2024.⁹ Figure A.4 reports analogous ACS results using the human coded exposure measure, which yields estimates closer to zero.

One potential concern is that post 2020 recovery dynamics could be correlated with exposure, particularly if reopening and demand shifts affected different artistic occupations unevenly. Figure A.5 addresses this by estimating a triple difference specification, plotting coefficients on artist by year by exposure interactions. These estimates reflect changes in the exposure gradient for artists relative to non-artists over time and show little evidence of differential pre trends, supporting the interpretation of the ACS patterns as changes that coincide with the diffusion of generative AI.

Figure A.6 in the Online Appendix reports analogous event study estimates using the county level AI exposure measure in Felten et al. (2021). This measure differs conceptually and temporally from the Eloundou et al. (2024) index: it reflects broader, pre generative AI notions of automation potential combined with local industrial composition rather than task exposure to large language models. The resulting estimates are negative in 2023 and 2024 but exhibit nonzero pre trends, which weakens their credibility as evidence of a discrete generative AI shock.

Table 1 reports heterogeneity by education, age, and degree field. The 2023 earnings coefficients are similar for college and non college respondents, while effects attenuate for non college respondents in 2024. Patterns by age are similar: both groups exhibit positive earnings coefficients

⁹Effects are similar when using log hourly wages as the outcome. Table A.5 reports the full ACS estimates with demographic controls. Figure A.1 replicates the analysis using 2017 as the base year.

in 2023, but effects are closer to zero in 2024 for younger workers. Respondents with a non fine arts college degree show a stronger earnings gradient with exposure in 2023 and 2024 than those with a fine arts degree, consistent with evidence that business or adjacent training can mitigate artists' earnings disadvantages.¹⁰ Table 2 reports analogous results for hours worked.

As a final robustness exercise, I implement a matched triple difference design within the two digit SOC 27 group. Artists and non-artists within SOC 27 can differ systematically in observable characteristics related to earnings and hours, and those differences may vary over time. To improve comparability and reduce functional form dependence and remaining selection effects, I use coarsened exact matching (CEM) (Iacus et al., 2011, 2012) to reweight the sample so that the distribution of basic demographics is balanced across artists and non-artists before estimating the event study. Specifically, I match exactly on five year age bins from 25 to 65 and on three education categories (high school or less, some college or associate degree, and college or more). Figure A.7 in the Online Appendix reports highly similar patterns to the baseline estimates.

These patterns are consistent with several competing channels. First, an augmentation or task reorganization channel predicts increases in effective throughput that may show up initially as higher labor input, especially if output is constrained by deadlines and client demand. Second, a scarce complementarity channel predicts that gains are concentrated among workers with characteristics that facilitate effective use, such as higher education, greater experience, or specific occupational niches. Third, a demand shock channel predicts that the exposure measure proxies for creative segments that were already expanding for reasons unrelated to AI, which could generate similar reduced form patterns. Finally, an efficiency wage or retention channel predicts wage responses that are not primarily mediated by hours and are linked to turnover or outside option

¹⁰See Makridis and Kuuskoski (2024).

considerations. The empirical design constrains these interpretations in several ways. Pre-trend diagnostics limit the scope for pre-existing differential trends. Industry and geography controls, along with within occupation comparisons where feasible, reduce the concern that exposure is simply a proxy for sectoral demand shifts. Decompositions that condition earnings on labor supply measures clarify whether earnings responses reflect pay versus hours. Complementary evidence on adoption and reported use cases provides external validation that generative AI is being used in ways that plausibly affect upstream creative tasks. Nonetheless, the next section examines a candidate mechanism behind these effects, namely the adoption of AI by artists and their use.

5 Understanding the Mechanisms

Given the slightly positive effects of genAI exposure on labor market outcomes, I turn to the Gallup Workplace Panel to understand the mechanisms. In particular, I exploit the panel structure of the data to run regressions of the form:

$$y_{it} = \gamma AI_{it} + \phi Artist_{it} + \xi(AI_{it} \times Artist_{it}) + \zeta X_{it} + \alpha_i + \lambda_t + \varepsilon_{it} \quad (3)$$

where y denotes our employee-level outcome of interest (e.g., burnout, job satisfaction), AI denotes a vector of indicators on AI utilization, $Artist$ denotes an indicator for being an artist, X is a vector for individual covariates as before, and α_i are individual and λ_t are year fixed effects. Standard errors are clustered at the individual-level and observations are weighted.

The primary concern in Equation 1 is that unobserved heterogeneity at the individual-level is correlated with AI adoption and/or sorting into different industries or occupations. To address these concerns, I not only include a wide array of demographic factors, but also introduce person

fixed effects, exploiting within-person variation across different employers. Unfortunately, the within-person variation is limited since the data on AI only began to be collected annually in 2023 and turned quarterly in 2025. Moreover, the sample of artists is small. While these factors lead to less statistically significant results, the coefficient estimates are economically meaningful.

Table 3 suggests that the mechanism is less about broad improvements in worker well-being from AI use, and more about who adopts AI (selection) and how heavy use may differentially map into artists' experiences. In pooled specifications that condition on demographics but omit person and year fixed effects, frequent AI utilization (“often”) is positively associated with job satisfaction (≈ 0.094 sd) and, simultaneously, with higher burnout (≈ 0.063 – 0.066 sd) relative to the “never/rare” group. Occasional use (“sometimes”) is also positively associated with job satisfaction (≈ 0.088 – 0.092 sd), but its association with burnout is economically small and statistically indistinguishable from zero (≈ 0.011 – 0.013 sd). These cross-sectional gradients are consistent with AI adoption being more common in roles that are more engaging and, for heavier users, also more demanding, so the raw associations need not reflect within-worker changes in well-being.

Once person and year fixed effects are introduced—shifting identification toward within-person changes in AI use over time—the main effects attenuate sharply and lose statistical significance for both outcomes (e.g., “often” falls to ≈ 0.031 – 0.033 sd for job satisfaction and ≈ 0.034 – 0.036 sd for burnout, while “sometimes” is near zero for both). The pattern reinforces the concern about unobserved heterogeneity correlated with AI adoption: much of the pooled association appears to be driven by time-invariant individual factors and common time shocks rather than within-person changes. The most salient within-person heterogeneity emerges when artists are defined at the industry level. In the fixed-effects specification, the interaction between frequent AI use and being an industry-defined artist is negative and statistically significant for job satisfaction (≈ -0.194 sd).

Given the small and imprecise main effect of “often” in that specification (≈ 0.033 sd), the implied net association for industry-defined artists is economically meaningful and negative (≈ -0.16 sd). By contrast, the analogous interactions for “sometimes” are close to zero and imprecisely estimated. One interpretation is that intensive AI use among industry-defined artists is more likely to coincide with changes in the content, process, or autonomy of creative work—channels that plausibly erode satisfaction even if AI raises measured productivity—whereas more episodic use does not.

For burnout, the interaction patterns are more nuanced and depend on how artists are classified. Using the occupation-based definition (SOC 27), the interaction between frequent AI use and artist status is negative and statistically significant in the fixed-effects specification (≈ -0.194 sd), implying that frequent AI adoption is associated with a meaningful decline in burnout for occupation-defined artists (net effect $\approx 0.036 - 0.194 \approx -0.16$ sd). In contrast, the industry-based interaction terms for burnout are imprecise and not statistically distinguishable from zero, and the “sometimes” interactions are likewise small and noisy across both definitions. Viewed in light with the attenuation of the main effects under fixed effects, the evidence points away from broad within-person well-being gains from AI use in the short 2023–2025 panel, and instead toward definition-dependent artist heterogeneity: intensive AI use appears most plausibly linked to lower satisfaction for industry-defined artists, but to lower burnout for occupation-defined artists.

Gallup asks respondents to explain how they are using AI, including in potentially overlapping categories. Figure 3 documents these results for the pooled sample and artists (occupation-based definition). Two patterns stand out. First, AI use is disproportionately oriented toward ideation in artistic work: the modal use-case for artists is *generating ideas*, and its share is markedly higher than in the pooled sample. Artists also report heavier use for *automating basic tasks* and, to a lesser extent, *learning new things* and *collaborating with others*, suggesting that AI complements both

upstream creative exploration and supporting workflows around production and coordination.

Second, relative to the pooled workforce, artists are less likely to use AI for operational or externally facing tasks. In particular, the artist distribution places less weight on *interacting with customers* and *making predictions*, and it is also lower for *setup/operate/monitor devices*. By contrast, *consolidating information* is common in both groups and sits among the top applications overall, indicating that AI’s value for summarization and synthesis generalizes beyond creative occupations. In sum, the results show that AI adoption among artists is skewed toward generative and creative-support functions rather than prediction or customer-facing applications.

6 Discussion

Technological disruptions have long jolted creative industries. A famous 19th-century example is the invention of photography, which contemporaries feared would make painters obsolete. Upon the debut of the daguerreotype in 1839, critics proclaimed it “the end of art,” predicting that painting would become “dead” or “obsolete” in the face of the camera. Indeed, portrait painters saw their commissions collapse as photographs could capture realism more cheaply. However, painting did not disappear. Instead, artists gravitated toward new styles (e.g., impressionism) that emphasized creativity over replication. Similarly, the digital revolution in music around the turn of the 21st century disrupted creative livelihoods and IP models. The rise of MP3 compression and peer-to-peer sharing (e.g. Napster) led to a “dramatic decline in revenues from recorded music”, with U.S. sales plunging from \$14.6 billion in 1999 to around \$9 billion by 2008.” This collapse in revenue profoundly affected musicians, songwriters, and studios, raising alarms about the viability of music careers. But the industry eventually adjusted: artists shifted focus

to live performances and new distribution models (such as streaming services), and overall the quantity of new music and artists continued to grow (Waldfoegel, 2017).¹¹ These historical cases illustrate a common pattern: new technology often displaces certain creative roles and undermines older business models (at least temporarily), but opens up new creative frontiers and eventually generates different kinds of jobs. Is generative AI (AI) different for the creative economy?

The results in this paper point to a familiar tension in the economics of technological change: generative AI appears to operate less like a discrete shock to workers’ realized earnings and more like a reallocation force that reshapes tasks, workflows, and the organization of production. In line with recent task-based frameworks, even when new technologies automate parts of a job, they can simultaneously expand complementary tasks and spur new categories of work (Acemoglu and Restrepo, 2019; Autor et al., 2024). That perspective helps reconcile why I find little evidence of immediate earnings declines for artists despite high measured exposure of many creative tasks to LLM capabilities (Eloundou et al., 2024; Maslej et al., 2024). In the short run, the dominant margin may be compositional and organizational—who adopts, in what settings, and for which tasks—rather than a one-for-one substitution of AI for labor.

A second interpretation is that the creative economy is experiencing a “productivity J-curve” dynamic: early adoption may raise measured output and hours without quickly translating into higher average earnings, particularly when complementary investments in intangible capital (new workflows, distribution channels, contract forms, and IP enforcement) lag behind (Brynjolfsson et al., 2021). This is consistent with emerging macro- and micro-evidence suggesting modest

¹¹Another parallel comes from film and animation. The introduction of computer-generated imagery (CGI) in the 1990s quickly superseded many practical special effects techniques that had employed craftspeople in model-making, matte painting, and puppetry. Traditional special effects artists had to retrain or find their skills in lower demand as studios embraced digital production for greater control and lower costs. At the same time, CGI unlocked new creative possibilities on screen and created an entire new occupation of digital visual effects artists.

(Humlum and Vestergaard, 2025) to somewhat positive (Johnston and Makridis, 2025) near-term labor market effects from LLM diffusion even when exposure measures indicate substantial task coverage. In creative settings, the gap between technical feasibility and economic impact may be especially wide because value creation depends on taste, differentiation, reputation, and more—features that do not adjust instantly to a new general-purpose technology (Waldfogel, 2017).

The Gallup panel evidence further underscores that the key adjustments may occur through within-job reorganization and changes in the experience of work, not only through wages and employment. In pooled cross-sections, frequent AI use correlates with higher job satisfaction and higher burnout, patterns consistent with selection into demanding but engaging roles. With person and year fixed effects, these associations attenuate substantially, suggesting that time-invariant worker heterogeneity and common shocks account for much of the raw gradient. Nonetheless, the remaining heterogeneity for artists hints at a deeper mechanism: intensive use may change the content and autonomy of creative work in ways that can depress satisfaction for some artist definitions even when it does not raise burnout on average. This aligns with arguments that generative AI should be understood as a new medium that reorganizes creative production—lowering some barriers to entry while raising new frictions around authorship, originality, and control (Epstein and Hertzmann, 2023; Zhou and Lee, 2024). More broadly, it resonates with long-standing definitional and institutional complexities in identifying, organizing, and representing “artists” as a labor market group (Skaggs and Aparicio, 2023).

Finally, these findings speak to an important baseline in cultural labor markets: artists have historically faced earnings penalties and heightened income risk even absent major automation shocks (Filer, 1986; Alper and Wassall, 2006; Makridis, 2025c). Generative AI therefore enters a sector already characterized by fragile career ladders and financially constrained institutions

(Makridis, 2025b). If AI primarily expands output capacity while novelty and differentiation become harder to sustain, market structure may tilt toward concentration and superstar dynamics in the sense of Rosen, because scalable tools can increase the payoff to reputation, distribution, and catalog depth even as they lower marginal production costs (Acemoglu, 2021). Conversely, if adoption lowers fixed costs of experimentation and distribution, it could enlarge creative participation and demand, echoing the historical pattern whereby digitization ultimately expanded cultural variety even after disrupting legacy revenue models (Waldfogel, 2017). Which force dominates is an empirical question that likely varies by medium and platform governance, including how recommendation systems allocate attention and how rights are enforced. Distinguishing between these channels will require longer-run data and sharper measurement of task reallocation within creative workflows, the contractual and IP environment, and the complementary skills that allow artists to appropriate returns from new tools (Bakhshi et al., 2015; Autor et al., 2024).

These considerations also suggest several near-term policy margins. If generative tools primarily reallocate tasks and raise the returns to complementary skills, a pragmatic response is not blanket restriction but targeted support for capability building, such as short-cycle training grants for artists and small creative firms focused on workflow integration, contracting, and business practices that preserve attribution and bargaining power. In parallel, disclosure and provenance standards for AI-assisted outputs could reduce consumer confusion and support market segmentation, allowing creators to credibly signal human-made versus AI-assisted production when that attribute is valued. These interventions are consistent with the paper’s results in that they aim to shape how adoption translates into economic returns, rather than assuming a one-for-one substitution of labor, which would only happen in extreme circumstances.

It is important to remember that creative labor markets are mediated as much by institutions

of appropriation as by the underlying task technology. As large language models are trained on vast corpora of digitized creative content, the absence of enforceable, granular property rights risks a commons problem in which works are consumed without clear authorization or remuneration, weakening incentives to invest in high-quality creation and verification. A plausible way forward is a hybrid, polycentric governance architecture that couples on-chain rights registration and automated licensing/royalty distribution with off-chain legal recognition and adjudication (Makridis and Ammons, 2025). Concretely, creators (or collective rights organizations) could mint standardized licenses for works and datasets—encoding permitted uses for training and downstream generation, attribution requirements, and pricing schedules—and route usage-based micropayments through smart contracts, while contested claims are resolved via specialized arbitration mechanisms (e.g., tokenized registries with dispute-resolution layers) and, when necessary, courts. This reframes policy away from restrictions and toward scalable provenance and compensation: a widely adopted registry linking content to verified creators and machine-readable licensing terms can preserve experimentation with generative tools while mitigating uncompensated substitution and related welfare concerns (Epstein and Hertzmann, 2023; Zhou and Lee, 2024; Acemoglu, 2021).

Generative AI also raises a place based policy question. Many creative labor markets remain geographically concentrated, in part because cities and designated cultural districts bundle amenities, audiences, institutions, and dense professional networks that support matching and learning. Evidence on culture amenities shows that local increases in arts activity are capitalized into housing values (Makridis, 2025a). Moreover, Makridis (2025a) conducts an event study showing that the designation of cultural districts at the county-level is associated with house price appreciation. Nevertheless, the diffusion of generative AI tools can make some task components less tied to local production ecosystems, particularly those that are digitally mediated, such as drafting, ideation,

editing, and some forms of post production. This creates the possibility of partial geographic decoupling, where creative hubs retain advantages for live, collaborative, and network intensive production, while a larger share of upstream and support tasks can be performed remotely. The net effect is therefore ambiguous and likely heterogeneous across artistic subfields, with implications for whether cultural district designations and production subsidies primarily preserve local employment, shift toward attracting complementary in person production, or increasingly function as demand side amenities that sustain local cultural consumption.

Several limitations are worth noting. First, the key treatment intensity is an occupation level, task based exposure index from Eloundou et al.; it is well suited to capturing exposure of text and planning intensive tasks, but it maps less directly to embodied and relational components of artistic work such as live performance, rehearsal, and improvisation, and average effects may therefore mask heterogeneity across artistic subfields. Second, the Gallup Workplace Panel provides useful evidence on adoption and use cases, but a larger creative worker sample and additional waves are needed to characterize longer run diffusion and adjustment. Third, while the design includes event study diagnostics and robustness checks, the estimates cannot fully rule out contemporaneous demand or institutional shocks that are correlated with measured exposure in the post 2022 period.

7 Conclusion

The historical record suggests that technological disruptions often unsettle creative industries, initially displacing certain roles and reshaping business models before ultimately fostering adaptation and new opportunities. From photography's impact on painting to digital compression's effect on music distribution, artists have repeatedly navigated the tension between automation and

creativity. Generative AI presents a similar inflection point. This paper evaluates how the rapid diffusion of large language models (LLMs) is translating into detectable labor-market harm for artists, and then uses panel evidence to shed light on mechanisms operating inside the workplace.

Using state–occupation data from the Occupational Employment and Wage Statistics (OEWS), the event-study evidence shows no sharp post-2020 break in the relationship between occupational generative-AI exposure and earnings or employment outcomes for artistic occupations. Pre-trends are close to zero, and post-2020 coefficients are generally small and imprecisely estimated, with confidence intervals that overlap zero across outcomes. If anything, the level estimates for earnings drift modestly upward after 2020, while employment estimates remain statistically indistinguishable from zero. Growth-rate results are more mixed—earnings growth is modestly positive in several post-2020 years, whereas employment growth is more often negative—but overall the OEWS evidence does not support large, contemporaneous declines in earnings or employment among more-exposed artistic occupations in the short run.

Individual-level evidence from the American Community Survey (ACS) complements these findings and helps address the OEWS limitation that establishment-based data miss many freelance arrangements common in the arts. The ACS event studies similarly exhibit near-zero pre-trends. Post-2020, exposure is associated with a modest increase in earnings in 2023 that attenuates by 2024, alongside a clearer positive association with total hours beginning in 2022 and persisting through 2024. The ACS patterns are more consistent with work intensification and task reallocation than with immediate wage erosion. Nonetheless, the short window and the imprecision of several estimates imply that medium-run dislocation remains an open empirical question, especially if adoption reshapes contracting, pricing, or demand in ways that unfold slowly.

To interpret these reduced-form patterns, I draw on the Gallup Workplace Panel to measure

AI utilization directly and examine within-worker associations with well-being. Artists report meaningfully higher AI adoption than the pooled workforce, especially frequent use, and their reported use-cases are disproportionately oriented toward ideation and creative-support functions (e.g., generating ideas, automating basic tasks, learning new things, and collaboration), rather than prediction, customer-facing applications, or operating devices. In pooled regressions, AI use is positively correlated with job satisfaction (and, for frequent users, burnout), consistent with selection into roles and workplaces that are simultaneously engaging and demanding. However, once person and year fixed effects are introduced, the main effects attenuate sharply, underscoring the importance of unobserved heterogeneity in AI adoption. The remaining heterogeneity is definition-dependent: intensive AI use is most plausibly linked to lower job satisfaction among industry-defined artists, while for occupation-defined artists it is associated with lower burnout in the fixed-effects specification. Although within-person variation is limited in the early years of measurement and the artist sample is modest, these results point to a mechanism in which AI is already being integrated into creative workflows, with nuanced implications for worker experiences.

These findings raise several priorities for future research. [Autor et al. \(2024\)](#) emphasize that whether new work emerges depends on whether innovations complement occupational outputs and coincide with demand shifts that raise occupational demand, whereas technologies that automate tasks or reduce demand can slow the emergence of new work. Against a backdrop of financial fragility for artists and arts institutions ([Makridis, 2025b,c](#)), understanding the demand-side response to generative AI—and the institutional environment governing appropriation, contracting, and distribution—will likely be central to assessing longer-run effects. A key next step is therefore to combine longer panels with richer measures of task content and creative production to distinguish productivity gains from work intensification, and to identify the margins through which

artists do (or do not) capture the returns to generative tools.

References

- Acemoglu, D. (2021). Harms of ai. Working Paper 29247, National Bureau of Economic Research.
- Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30.
- Acemoglu, D. and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, 128(6):2188–2244.
- Alper, N. O. and Wassall, G. H. (2006). Artists' careers and their labor markets. *Handbook of the Economics of Art and Culture*, 1.
- Autor, D. H., Chin, C., Salomons, A., and Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. *The Quarterly Journal of Economics*, 139(3):1399–1465.
- Bakhshi, H., Frey, C. B., and Osborne, M. (2015). Creativity vs. Robots: The Creative Economy and the Future of Employment. Technical report, Nesta, London.
- Brittain, B. (2023). Getty images lawsuit says stability ai misused photos to train ai. *Reuters*.
- Brynjolfsson, E., Chandar, B., and Chen, R. (2025). Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence. Technical report, Stanford Digital Economy Lab.
- Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? In *AEA Papers and Proceedings*, volume 108, pages 43–47.

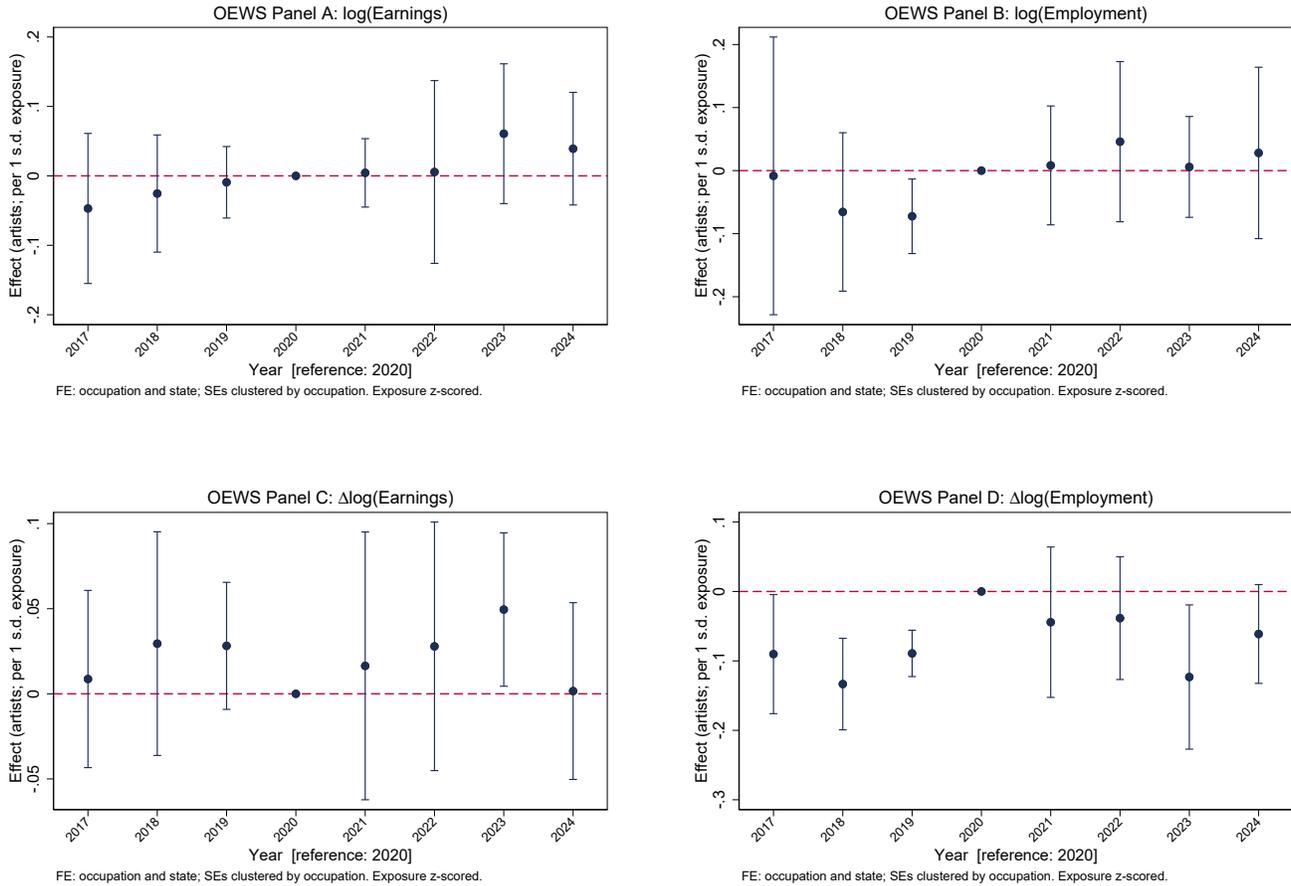
- Brynjolfsson, E., Rock, D., and Syverson, C. (2021). The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372.
- Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308.
- Epstein, Z. and Hertzmann, A. (2023). Art and the science of generative ai. *Science*, 380:1110–1111.
- Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 47(12):2195–2217.
- Filer, R. K. (1986). The “starving artist”- myth or reality? earnings of artists in the united states. *Journal of Political Economy*, 94(1):56–75.
- Frey, C. B. and Osborne, M. A. (2017). The Future of Employment: How Susceptible are Jobs to Computerisation? *Technological Forecasting and Social Change*, 114:254–280.
- Humlum, A. and Vestergaard, E. (2025). Large language models, small labor market effects. *BFI working paper*.
- Iacus, S. M., King, G., and Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, 106:345–361.
- Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20:1–24.

- Johnston, A. and Makridis, C. A. (2025). The labor market effects of generative ai: A difference-in-differences analysis of ai exposure. *SSRN working paper*.
- Makridis, C. A. (2025a). Capitalization effects for culture amenities. *SSRN working paper*.
- Makridis, C. A. (2025b). Financial headwinds in the performing arts: Evidence from united states opera, 2005–2023. *Journal of Arts Management, Law, and Society*, 49:1–20.
- Makridis, C. A. (2025c). The labor market returns to artistic occupations: Evidence from the united states. *Journal of Cultural Economics*, 49:1–21.
- Makridis, C. A. (2025d). Organizational transmission of ai: Managerial influence on generative ai adoption. *SSRN working paper*.
- Makridis, C. A. and Ammons, J. D. (2025). Governing the large language model commons: Using digital assets to endow intellectual property rights. *Journal of Institutional Economics*, 21(e19):1–20.
- Makridis, C. A. and Kuuskoski, J. (2024). Narrowing the gap: Implications of arts entrepreneurship curricula on artist labor market outcomes. *Artivate*, 12(1).
- Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J., Shoham, Y., Wald, R., and Clark, J. (2024). The ai index annual 2024 report. *Institute for Human-Centered AI, Stanford University, Stanford, CA*.
- McCouat, P. (2015). Early influences of photography on art, part 1.

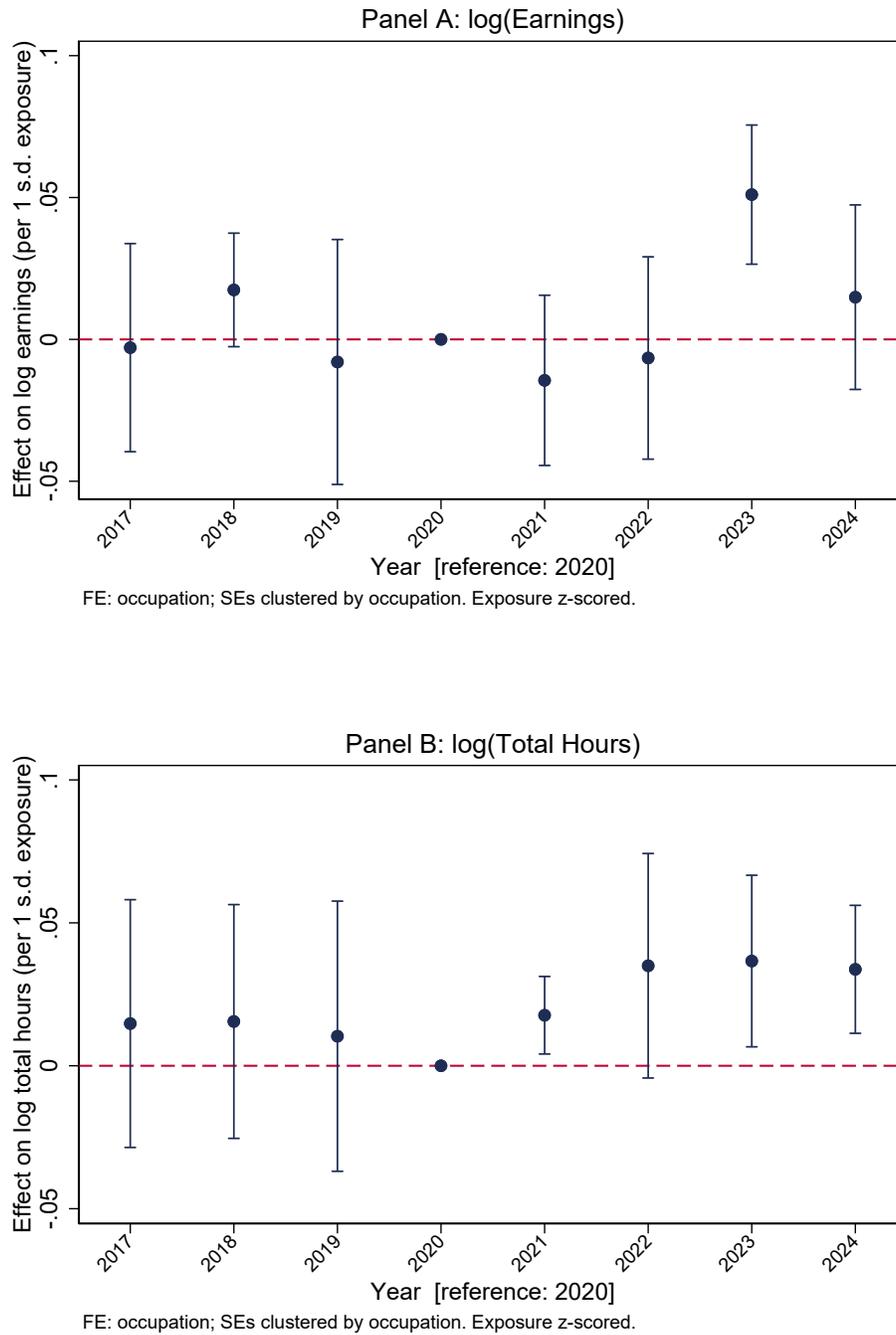
- Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., and Sobek, M. (2024). IPUMS USA: Version 11.0 [dataset]. *Minneapolis, MN: IPUMS*.
- Skaggs, R. and Aparicio, T. (2023). Workers and work in the arts: Definitional challenges and approaches to collective action among arts and creative workers. *Work and Occupations*, 50(3):436–444.
- Waldfogel, J. (2017). How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives*, 31(3):195–214.
- Zhou, E. and Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, 3(3):pgae052.

Tables and Figures

Figure 1: Dynamic Effect of Generative AI Exposure on Labor Market Outcomes



Notes.—Sources: Occupational Employment and Wage Statistics (2017-2024) and [Eloundou et al. \(2024\)](#). The figure reports the coefficients associated with regressions of year-to-year earnings and employment growth (with earnings deflated with the 2017 personal consumption expenditure index) on interactions of year fixed effects and a z-score of large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, conditional on SOC and year fixed effects (normalized to 2020). From [Eloundou et al. \(2024\)](#), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the LLM-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Employment and earnings growth is winsorized at the top and bottom percentile. Standard errors are clustered at the occupational-level.

Figure 2: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked

Notes.—Sources: [Eloundou et al. \(2024\)](#), American Community Survey (2017-2024). The figure reports the coefficients associated with regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z -score of β displacement risk of at least 10% of tasks (LLM-coded) from [Eloundou et al. \(2024\)](#), interacted with year dummies (normalized to 2020), conditional on SOC and year fixed effects. Demographic factors are included as controls: male, marital status, a quadratic in age, race (White, Black), and a quadratic in years of schooling. Standard errors are clustered at the occupation-level and observations are weighted by the sample weights.

Table 1: Heterogeneity in Large Language Model Exposure on Individual Earnings

	log(Annual Earnings)					
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure \times Year = 2017	.033** [.012]	-.079 [.118]	-.027 [.029]	.011 [.046]	-.028 [.036]	.090* [.040]
Exposure \times Year = 2018	.144** [.043]	-.062 [.106]	.019 [.062]	.079* [.031]	.198* [.076]	.092* [.039]
Exposure \times Year = 2019	.012 [.044]	-.072* [.030]	-.026 [.062]	-.024 [.050]	-.099 [.062]	.107 [.053]
Exposure \times Year = 2021	-.036 [.046]	-.023 [.100]	-.093 [.098]	.001 [.049]	-.215*** [.046]	.137 [.135]
Exposure \times Year = 2022	-.041 [.033]	.010 [.121]	.002 [.014]	-.027 [.066]	-.117 [.074]	.058* [.021]
Exposure \times Year = 2023	.118*** [.022]	.191** [.059]	.152 [.094]	.154* [.061]	.063 [.053]	.181*** [.028]
Exposure \times Year = 2024	.078 [.055]	-.014 [.047]	.116** [.026]	.010 [.046]	.084 [.076]	.076 [.084]
R-squared	.14	.11	.12	.13	.13	.14
Sample Size	20143	8204	14048	14299	7850	12293
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Sample	College	Non-College	Over 40	Under 41	Fine Arts	Non-Fine Arts

Notes.—Sources: American Community Survey (2017-2024) and Eloundou et al. (2024). The table reports the coefficients associated with regressions of log annual earnings (with earnings deflated with the 2017 personal consumption expenditure index) on interactions of year fixed effects and large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, separately for different groups (college v. non-college; over 40 v. under 41; with fine arts degree v. non fine arts degree but still college), conditional on SOC and year fixed effects (normalized to 2020). From Eloundou et al. (2024), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the LLM-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Standard errors are clustered at the occupation-level.

Table 2: Heterogeneity in Large Language Model Exposure on Individual Hours Worked

	log(Annual Hours)					
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure \times Year = 2017	.002 [.015]	.094 [.089]	-.026 [.073]	.089*** [.011]	.038 [.047]	-.028 [.026]
Exposure \times Year = 2018	.037 [.022]	.069 [.066]	.014 [.060]	.066* [.025]	.132 [.077]	-.042 [.056]
Exposure \times Year = 2019	.013 [.025]	.055 [.074]	.055 [.063]	.016 [.043]	.022 [.058]	.005 [.025]
Exposure \times Year = 2021	.004 [.020]	.126* [.048]	.026 [.069]	.067 [.032]	-.029 [.040]	.041 [.045]
Exposure \times Year = 2022	.072** [.018]	.145* [.065]	.037 [.051]	.142*** [.030]	.097** [.028]	.066*** [.012]
Exposure \times Year = 2023	.043** [.011]	.181** [.058]	.096** [.035]	.106** [.031]	.083 [.057]	.020 [.035]
Exposure \times Year = 2024	.044** [.014]	.166*** [.020]	.081 [.064]	.102*** [.015]	.093* [.039]	-.002 [.026]
R-squared	.06	.07	.07	.06	.06	.06
Sample Size	26300	13908	20542	19666	10825	15475
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Sample	College	Non-College	Over 40	Under 41	Fine Arts	Non-Fine Arts

Notes.—Sources: American Community Survey (2017-2024) and Eloundou et al. (2024). The table reports the coefficients associated with regressions of log annual hours worked on interactions of year fixed effects and large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, separately for different groups (college v. non-college; over 40 v. under 41; with fine arts degree v. non fine arts degree but still college), conditional on SOC and year fixed effects (normalized to 2020). From Eloundou et al. (2024), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the LLM-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Standard errors are clustered at the occupation-level.

Table 3: Effects of AI Utilization on Work Outcomes for Artists

Dep. var. =	Job Satisfaction				Burnout			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI Use - Often	.094***	.031	.094***	.033	.066***	.036	.063***	.034
	[.019]	[.024]	[.019]	[.024]	[.021]	[.032]	[.021]	[.032]
× Artist (occ)	.042	.098			-.045	-.194**		
	[.082]	[.096]			[.097]	[.082]		
× Artist (ind)			.073	-.037			.093	-.110
			[.084]	[.104]			[.091]	[.106]
AI Use - Sometimes	.092***	-.001	.088***	.002	.011	-.001	.013	-.001
	[.017]	[.019]	[.017]	[.019]	[.019]	[.021]	[.019]	[.021]
× Artist (occ)	-.085	.098			.051	-.000		
	[.078]	[.078]			[.103]	[.066]		
× Artist (ind)			.042	-.028			-.014	-.014
			[.074]	[.117]			[.082]	[.065]
R-squared	.02	.75	.02	.75	.05	.79	.05	.79
Sample Size	68382	56623	68382	56623	55793	44112	55793	44112
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Person FE	No	Yes	No	Yes	No	Yes	No	Yes
Year FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes.—Sources: Gallup Panel, 2023-2025. The table reports the coefficients associated with regressions of a z -score for job satisfaction and burnout on indicators of AI utilization, their interactions, and conditional on individual covariates and fixed effects. AI Use - Often is based on using AI daily or a few times a week; AI Use - Sometimes is based on using AI a few times a month or a few times a year; the normalization is relative to those who use it once a year, less than once a year, or never. Job satisfaction is similarly defined: “How satisfied are you with your place of employment as a place to work?” Burnout is a z -score to: “Please indicate how often each of the following statements is true of your primary job (i.e., the job where you spend the most time working). I feel burned out at work: (a) Always, (b) Very often, (c) Sometimes, (d) Rarely, (e) Never” from (5=always) to (1=never). I rely on two complementary approaches to classifying artists: an industry-based definition (`artind`, Arts/Design/Entertainment/Sports/Media) and an occupation-based definition (`artocc`, SOC = 27, Arts, Design, Entertainment, Sports, and Media Occupations). Demographics include: age, race (White, Black), full-time status, having children, and being married. The sample is restricted to full and part time workers, and the observations are weighted.

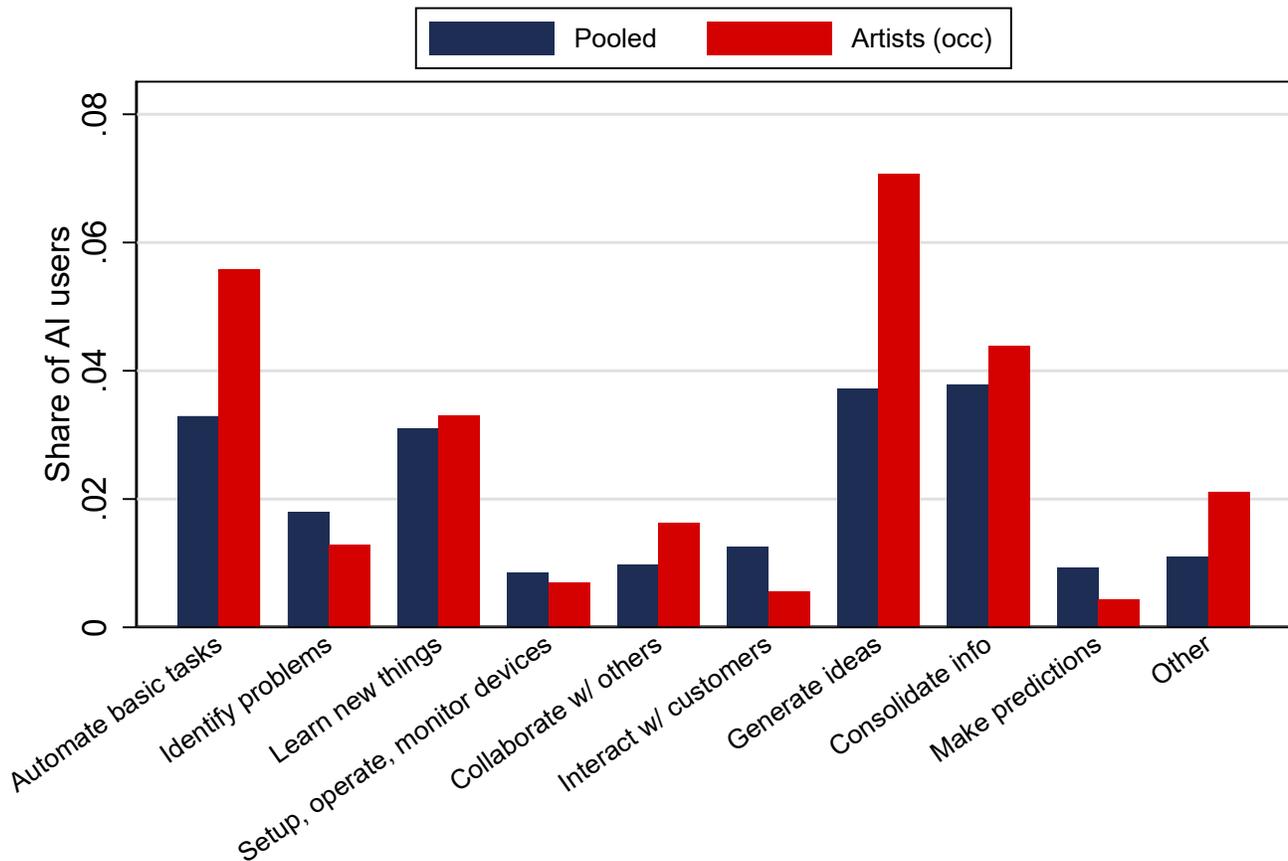


Figure 3: Composition of AI Use-cases by Artists and non-Artists

Notes.—Sources: Gallup Panel, 2024-2025. The figure plots the share of respondents across different AI applications distinguishing among the pooled sample and artists. AI Use - Often is based on using AI daily or a few times a week; AI Use - Sometimes is based on using AI a few times a month or a few times a year; the normalization is relative to those who use it once a year, less than once a year, or never. I measured artists based on an occupation-based definition (`artocc`, SOC = 27, Arts, Design, Entertainment, Sports, and Media Occupations). The sample is restricted to full and part time workers, and the observations are weighted.

A Online Appendix

A.1 Additional Materials from Eloundou et al. (2024)

Here, I summarize some of the most relevant background from Eloundou et al. (2024) that is contained in their Online Appendix, which is helpful for contextualizing my study.

Eloundou et al. introduce an exposure rubric to classify tasks by how large language models (LLMs) like GPT-4 could accelerate their completion without quality loss. Human experts (and a tuned GPT-4 prompt) labeled each task into one of several exposure levels. E0 (no exposure) means access to an LLM would not cut task time by at least half at equivalent output quality. E1 (direct exposure) indicates the task can be substantially accelerated (over 50% time saved) by GPT-4 alone (e.g. via ChatGPT or coding assistant) with no drop in quality. E2 (LLM-assisted exposure) applies to tasks that cannot be sped up enough by an LLM alone, but could be with “LLM + software,” i.e. if complementary tools or system integration (built on the LLM’s capabilities) are added. In other words, E2 tasks require some additional application (such as feeding the LLM longer documents, retrieving up-to-date information, or interacting with external data/services) to realize a 50% time reduction. A final category, E3, was defined for tasks that would become exposed given multimodal (image) capabilities: tasks where an LLM coupled with vision (reading/creating images) would significantly speed work (e.g. reading diagrams or generating designs). This E3 category was included in the rubric for completeness but represents capabilities beyond text; in analysis E3 cases were generally merged with E2 (since both require additional systems) or treated as unexposed under current text-only models.

Rubric design and task evaluation: The 50% time-saving threshold (with equal quality)

was chosen as a clear, conservative bar for “exposure.” It is somewhat arbitrary but meant to flag only substantial productivity gains. In practice, the authors note this high threshold roughly corresponded to cases where “an LLM or LLM-powered software could perform the core part of a task or nearly the entire task”. Annotators were instructed to assume a capable GPT-4-level model with a 2,000-word context window, knowledge up to 1 year old (unless provided), and usual digital tools (laptop, internet, etc.), but no specialized physical tools. Under these conditions, they judged each task’s exposure level per the rubric definitions. For example, a task requiring complex writing or coding might be E1 if GPT-4 can directly handle it, whereas a task involving proprietary lengthy documents or databases could be E2 (since a custom tool integrating the LLM with those data would be needed). Tasks heavily reliant on human physical skills or in-person interaction are E0 by definition. This rubric explicitly does not distinguish between the task being automated vs. merely assisted – it measures potential technical efficiency gain, not whether the LLM replaces the worker or augments them. In fact, the authors emphasize that exposure is a proxy for potential impact under ideal conditions; it does not guarantee real productivity or automation outcomes in practice (factors like adoption, regulation, and social acceptance also play roles).

From task labels to exposure measures: Each occupation’s exposure index is then computed by aggregating its task labels. The authors consider three versions of the exposure measure: (i) Direct exposure (E_1) – the share of tasks labeled E1 (LLM alone) in that occupation; (ii) Partial exposure ($E_{1+0.5}$) – a weighted share counting E1 tasks fully and E2 tasks at half-weight (reflecting that E2 tasks require extra investment/tools); and (iii) Full potential exposure (E_{1+2}) – the share of tasks labeled E1 or E2, counting both as fully exposed (an upper bound assuming complementary software is developed for all E2 tasks). In all cases E0 tasks count as 0 exposure. (E3 tasks are generally counted as E2 if considering future multimodal tools, but are otherwise

excluded from current model exposure calculations.) Tasks were aggregated up from ONET’s Detailed Work Activities to broader task statements and then to occupations. To weight the importance of tasks, the main specification uses ONET’s designation of “core” vs. “supplemental” tasks: core tasks are given double weight relative to supplemental ones (so that core duties collectively get higher influence on the occupation’s exposure). All task weights are normalized to sum to 1 per occupation. The authors also examined alternative weighting schemes – e.g. treating all tasks equally – and found the overall exposure results to be robust (figures remained similar). Each of the exposure metrics E_1 , $E_{1+0.5}$, E_{1+2} can thus be reported as a fraction of tasks (0 to 1) exposed under different assumptions. By design, $E_{1+0.5}$ (their “preferred” measure) lies between the direct-only exposure and the full hypothetical exposure with complementary tools.

Direct vs. augmented exposure: Conceptually, an “exposed” task is one where the model can significantly speed up task completion at equal quality – either directly (E1) or via augmentation through new tools (E2). An E1 task implies an individual worker using ChatGPT/GPT-4 could at least double their productivity on that task. An E2 task implies that, while the model alone isn’t enough, a software application leveraging the model (for example, a bespoke data retrieval or document analysis tool) could unlock comparable efficiency gains. In other words, E1 captures immediate LLM capabilities out-of-the-box, whereas E2 captures latent LLM capabilities that need integration into workflows. This distinction also corresponds to direct labor substitution versus complementary augmentation in an occupational context – though the exposure index deliberately does not label them as such, since both could affect labor demand. The rubric’s “equivalent quality” clause ensures that speed-ups counted as exposure do not come at the cost of accuracy or standards, underscoring that the model must perform the task competently. Finally, because the exposure index is rooted in current technical feasibility, the authors caution it should

be interpreted as a relative measure of impact potential across occupations, not a literal projection of automation. For instance, an occupation with exposure 0.6 should be viewed as far more exposed than one at 0.1, but not necessarily that 60% of its tasks will be automated in practice. The exposure index simply quantifies how much of the job’s content could be significantly affected by GPT-like technology under ideal conditions.

Validation and reliability: The exposure labeling process involved both domain experts and GPT-4 itself. The authors first had humans (familiar with LLM capabilities) label a large sample of tasks and DWAs, then prompted an early GPT-4 model to classify tasks according to the same rubric. They iteratively refined GPT-4’s prompt with examples to improve agreement with human labels. In the end, GPT-4’s task classifications closely mirrored the human annotations: at the task level, the model and human labelers agreed on the exact E0/E1/E2 category roughly 80% of the time. At the occupation level (after aggregation), the exposure scores from GPT-4 and from humans are highly correlated (Pearson $r \approx 0.65$ for the full exposure measure). Figure 2 of the supplement plots GPT-4 vs. human occupation exposure and shows a tight positive relationship along the diagonal. (Notably, human raters tended to assign slightly higher exposure than GPT-4 for the most exposed occupations, suggesting humans were a bit more optimistic about some top-end cases.) The authors treat the human-based exposure as the primary measure in their analysis, but the strong concordance with GPT-4’s independent assessment provides reassurance that the rubric was applied consistently. They also cross-checked their exposure index against prior automation indices and found it aligns well with measures of software and AI-exposure from past work, while showing expected divergence from physical automation risk measures (since manual tasks are largely E0). In sum, the exposure index methodology combines expert judgment with LLM assistance to efficiently quantify how much of each job could be impacted by GPT-

like AI, distinguishing between tasks directly handled by the AI versus those requiring supportive technology. This yields a nuanced measure of “LLM impact potential” that can be used to compare occupations in terms of their susceptibility to GPT-driven productivity enhancement.

Exposure Rubric Taxonomy (E0, E1, E2, E3) and Examples: In the supplemental appendix, Eloundou et al. provide a detailed rubric for categorizing task exposure to GPT capabilities. The rubric defines four exposure levels:

- **E0 – No Exposure:** Tasks for which using an LLM (via a chat interface or similar) does not meaningfully speed up completion. For instance, tasks that are predominantly physical or require in-person human interaction are E0 by default. Example: “Adjust, clean, or repair equipment to correct defects” – this hands-on task was labeled E0 because a model cannot perform physical adjustments; having an LLM has no impact on the time needed.
- **E1 – Direct LLM Exposure:** Tasks that can be completed at least twice as fast with direct access to an LLM alone (while maintaining equivalent quality). These typically involve information processing, writing, or coding that the model can handle out-of-the-box. Examples: drafting text or code, translating or summarizing documents, generating content or answers from provided information. One example given is a research scientist’s task of applying theoretical knowledge to create new technology – labeled E1 because GPT-4’s training likely includes vast theoretical knowledge that it can leverage to suggest innovations via text, thus significantly accelerating that task.
- **E2 – LLM Exposure via Tools:** Tasks for which an LLM alone isn’t enough to halve the completion time, but could be if coupled with additional software or system integration. These are cases where one can “easily imagine” a specialized application built on the LLM

that would make the task much more efficient. Such software might handle long documents, access up-to-date data, or interface with other systems that the raw model can't. The appendix lists examples like a customer service tool that suggests responses using GPT-4, a legal research assistant that combs through case law, or a grading system that uses GPT to score essays with a given rubric. In short, E2 tasks are latent opportunities for automation – the model provides the core language/coding ability, but additional engineering is required to apply it to the task. Example: “Negotiate purchases or contracts” was marked E2 – while an LLM alone might not directly negotiate, one could imagine a system where each party’s statements are transcribed and the LLM helps draft agreements or suggest compromises. (This would require multiple users adopting the tool and trusting an AI mediator, so an integrated solution is needed.) Another example: “Prescribe medication...” (a doctor’s task) was labeled E2 – GPT-4 can propose likely diagnoses and write draft prescriptions or notes, but a human doctor must still review and finalize the decision. In these cases, the LLM can augment the task significantly, but cannot yet fully replace the human’s role without additional systems or authority.

- E3 – Exposure with Image Capabilities: Tasks that become exposed given an LLM that can process images (in addition to text). This category extends the rubric to multimodal scenarios. Raters assumed access to a hypothetical system that can “view, caption, and create images” (but not video, and not with fine-grained measurement accuracy) in conjunction with the LLM. A task was labeled E3 if with this vision-enabled AI, the task could be completed much faster (over 50% time saved). Typical examples include reading text from scanned PDFs, interpreting diagrams or photos, or generating/editing images per in-

structions. For instance, an administrative task like extracting data from handwritten forms could become quick if an AI could read the images. In the rubric, E3 was kept separate to capture tasks that need non-text capabilities. However, in the main analysis the authors combine E2 and E3 tasks together when estimating overall exposure. This is because both E2 and E3 represent tasks needing additional technology beyond a standalone LLM (whether that’s a database or an image processor). In effect, the reported “exposed” share of tasks often includes E3 tasks as part of the broader augmented exposure via complementary tools. (Where clarity is needed, the authors sometimes refer to exposure “with other generative tools” to encompass E3.) Notably, if one considers future multimodal GPT models, many E3 tasks might eventually become E2 or even E1 tasks.

Automation” Rubric (T0–T4) Design and Mapping: In addition to the exposure rubric, the authors experimented with a second classification in the appendix that explicitly gauges automation potential. This supplemental automation rubric assigns tasks to five levels (T0 through T4) reflecting the degree to which an LLM or LLM-driven system could perform the task without human assistance. The design intent here is to differentiate tasks that AI might fully take over from those where human involvement remains necessary (even if augmented by AI). In other words, this rubric tries to capture a spectrum from no automation to full automation, as a complement to the exposure measure. The mapping is defined as follows:

- T0 – No automation exposure: the task cannot be meaningfully automated by current GPT-like technology (essentially equivalent to E0 tasks). Score = 0.0.
- T1 – Low automation feasibility: a small portion of the task could be automated, but the majority still requires a human. Score = 0.25.

item T2 – Moderate automation feasibility: roughly half of the task could be done by an LLM or related tools, while the rest needs human input. Score = 0.5.

- T3 – High automation feasibility: an LLM could handle a large majority of the task, with only minor human involvement needed. Score = 0.75.
- T4 – Full automation exposure: the task could in principle be fully automated by an LLM-based system (100% automated). Score = 1.0.

The authors’ appendix indicates that they coded $T0 = 0$ (no AI help possible) and $T4 = 1$ (entire task automatable), with the intermediate T1, T2, T3 mapped to 0.25, 0.5, 0.75 respectively. This rubric forced a finer-grained judgment on how completely AI could replace a human for each task. A T4 task would be one where GPT-4 (with any needed tools) could do everything that a human worker does in that task, whereas T1 might imply the AI can do a minor subtask or provide guidance but a human is still needed for most of it. The design intent was to capture labor-displacing potential more directly: T4 corresponds to potential full automation (displacement), while lower T’s indicate the task remains largely in the human realm (with AI only augmenting).

Of note, the “T-rubric” was a supplemental analysis; the paper’s main results rely on the E-rubric exposure measures. The automation rubric was not as extensively validated with human raters. However, the authors report that an early GPT-4 labeling using the T0–T4 scheme yielded patterns broadly consistent with the E0–E3 exposure labels. For instance, tasks labeled E1 (directly exposed) often correspond to high T-levels (T3 or T4, since if GPT can handle the core of a task, a large portion is automatable). Tasks labeled E2 might map to intermediate T-levels (partial automation), and E0 tasks would be T0 in most cases. The appendix notes that the GPT-4 “Automation Rating” using the T-rubric did not significantly correlate with Frey &

Osborne’s classic automation risk scores, highlighting again that many traditionally “automatable” routine jobs involve physical activities outside of LLM capabilities. This underscores that T0–T4 was capturing a different facet – focusing on AI’s ability to fully substitute for a human in a task – whereas the exposure rubric allowed for AI as a tool (not necessarily a replacement). In summary, the T0–T4 automation rubric provided a quantitative scale (0 to 1) by which the authors could test alternate weightings or compare “full automation” potential, but it remains a secondary framework. The main text notes a high correlation between the exposure index and this automation-based index, suggesting that tasks heavily exposed to GPTs are also those most amenable to automation, even if the rubric definitions differ. (The exposure rubric was preferred for the study because it does not force a judgment on augmentation vs. substitution – it simply measures impact potential.)

Task Weighting, Aggregation, and Exposure Index Construction: The supplemental appendix details how task-level labels were aggregated into occupation-level exposure scores. Each occupation in O*NET comes with a list of tasks, some marked “core” (essential duties) and others “supplemental.” The authors leveraged these distinctions in their weighting scheme. By default, core tasks are given double the weight of supplemental tasks when computing an occupation’s exposure index. In practical terms, they assign a weight of 2 to core task statements and 1 to supplemental ones, then normalize so that the total weight across all tasks equals 1 for that occupation. This ensures the exposure score is more influenced by the primary job duties. (For example, if a core task is exposed but several minor supplemental tasks are not, the occupation would still register a high exposure because the core responsibilities are affected.) This was done to “emphasize tasks that are more critical to the occupation” while still acknowledging all tasks.

As a robustness check, [Eloundou et al. \(2024\)](#) also considered equal-weighting all tasks. In

fact, some of the figures in the supplement report results under equal weighting to demonstrate the difference. For instance, Figure 3 in the appendix shows the distribution of occupational exposure levels across the economy with tasks equally weighted. The general findings do not change qualitatively – giving each task equal importance yields a similar overall exposure pattern, though the absolute values shift slightly. (The main text notes that even with alternative aggregation methods, such as taking whichever label (human or GPT) was more “exposed” for a task, the concordance between human and model assessments remained high.) Ultimately, the paper’s headline results use the core-weighted scheme, but the supplement’s footnotes confirm that supplemental tasks were effectively counted as half a task for exposure summation. This weighting choice is also applied when aggregating to higher levels (like industry exposure): for example, Appendix Figure 6 and 7 in their paper display industry-level exposure where each occupation’s contribution is weighted by its employment share, and within each occupation the core vs. supplemental task weighting has been applied. The result is an employment-weighted exposure by industry, showing, say, that data-intensive industries have high exposure while manual industries have low exposure – a pattern consistent between human and GPT-4 labeling methods.

Another aspect of aggregation is how Detailed Work Activities (DWAs)—very granular descriptions—were handled. The appendix clarifies that human annotators initially labeled DWAs, which were then rolled up into the broader O*NET task statements for each occupation. In cases where multiple DWAs mapped to one task and had conflicting labels, the authors had to reconcile them. They experimented with taking the “maximally exposed” label versus majority vote, etc., and found overall exposure scores were not very sensitive to these choices. They report that even using a “maximum-match” approach (if either a human or GPT labeled a sub-task as exposed, count it as exposed) yielded similar agreement levels. To ensure accuracy, they did go back and

collect additional labels for certain task–occupation pairs where there was significant disagreement or ambiguity. (For example, annotators sometimes lacked context about how a task applied in a specific occupation, leading to initial mislabeling – Appendix Table 1 provides examples of such discrepancies and how they were resolved.) After these refinements, each occupation ended up with a set of task exposure labels that feed into the weighted averages described above.

Inter-Rater Agreement and Validation Checks: The supplemental materials also document various reliability and validation checks for the exposure annotations. A key focus is the agreement between human experts and GPT-4 as annotators. The authors present a comparison in Appendix Table 2, summarizing how often the labels coincided and how strongly the occupation-level scores correlated. For the direct exposure category (E1) alone, human vs. GPT-4 agreement was about 81% (i.e. they picked the same label E0/E1/E2 for four out of five tasks). When considering the full exposure measure (E1+E2 combined), the Pearson correlation between human-derived and GPT-derived occupation scores is around 0.65–0.70. Figure 2 in the appendix illustrates this with a scatterplot: each point is an occupation, plotting human exposure on one axis and GPT-4 exposure on the other. The points cluster tightly along the diagonal, indicating strong agreement overall. The figure also includes a binned trend line showing that at lower exposure levels the two methods align almost perfectly, while at the very highest exposure levels human raters gave slightly higher scores than GPT-4 did. The authors interpret this as GPT-4 being somewhat more conservative about labeling tasks as exposed in a few occupations, whereas humans may have been a bit more optimistic about GPT-4’s capabilities in those cases. Even so, there were no systematic divergences – the rank-ordering of occupations by exposure is very similar whether one uses human labels or GPT-4 labels.

Another validation mentioned is the concordance among human annotators themselves and the

steps taken to resolve disagreements. The humans involved were a small group of experts familiar with LLMs (including some of the authors and OpenAI annotators), but not necessarily domain experts in each occupation. As noted in their Appendix A.3, this led to some inconsistent labels when a task’s interpretation depended on domain context (e.g. understanding a task in a specific industry). Their Appendix Table 1 gives examples where initial human labels differed or seemed counter-intuitive, often due to annotators not knowing the occupation behind a task description. The authors addressed these by providing additional context or re-assigning those tasks for further review. They also tried aggregating at the task level in different ways (such as taking a consensus or even giving GPT-4 a chance to weigh in) and found the overall exposure metrics were robust. Ultimately, they obtained a refined set of human labels that served as the “ground truth” for the analysis, with GPT-4’s labels used mainly for comparison and scaling to the full task set. The high agreement with GPT-4’s independently generated labels provides some external validation that the rubric was clear and applied consistently.

Their appendix also compares the exposure measures with prior automation-related indices as a form of convergent validity. In Table 8 and Table 9 of the supplement, the authors regress their exposure scores against measures like Frey & Osborne’s (2017) automation probabilities, Webb’s (2020) AI/robotics exposure, Felten et al.’s (2018) AI suitability, and routine task content indices. The GPT exposure metric correlates positively with software and AI-based exposure measures from prior work, but is negatively correlated with physical automation risk and routine-manual work measures. This makes intuitive sense: the new metric identifies white-collar and cognitive tasks (often non-routine) as most exposed, whereas traditional automation risk was highest for routine manual jobs (which are E0 in the GPT rubric, since robots or machinery – not LLMs – would be needed to automate them). The supplement notes that their “automation rubric” (T0–T4) results

likewise did not align with Frey and Osborne (2017)’s probabilities reflecting that what GPT-4 can automate is a different subset of work than what physical automation can. These comparisons support the validity of the exposure index as capturing a novel axis of technological impact. The exposure scores also show plausible patterns across job categories: Appendix Table 10, for example, groups occupations by their required education level and reports mean exposure scores for each group. Occupations requiring only a high school diploma have relatively low exposure (e.g. human-labeled direct exposure 0.09 on average), whereas jobs requiring a bachelor’s or higher have much greater exposure (e.g. 0.23 for direct E1 exposure in bachelor’s-level occupations). In the highest education jobs (and highest wage brackets), a larger fraction of tasks are language- or knowledge-based, which explains the higher LLM exposure. This pattern – higher exposure in higher-education occupations – is borne out in Table 10 and is contrary to some earlier automation trends. It provides face validity that the rubric is indeed capturing the intuitive idea that GPTs most affect jobs heavy in information processing and coding, not manual labor.

Finally, the appendix discusses some limitations of their methodology, acknowledging the subjective element of the rubric and the forward-looking assumptions involved. The authors note that their annotators, while knowledgeable about GPT-4, were not diverse across all occupations, and some bias or misjudgment could occur. They suggest that involving domain experts for each occupation could further improve label accuracy in future work. They also reiterate that the exposure index measures technical capability under ideal conditions, not a prediction of immediate job loss – real-world adoption may be slower or face barriers (organizational, regulatory, ethical) that the rubric does not capture. Despite these caveats, the supplemental analyses (inter-rater checks, weighting robustness, and external comparisons) lend credibility to the exposure measures. In summary, the appendix material reinforces that the exposure index is a carefully constructed,

well-validated metric for the potential impact of GPT-level AI on job tasks, distinguishing between tasks directly handled by AI and those requiring complementary innovations to be affected. Figures and tables in the supplement substantiate these points: Figure 2 demonstrates the high human–GPT label agreement, Figure 3 highlights how complementary tools can vastly expand exposure, Figures 6 and 7 show broad industry exposure using the weighting scheme, and Table 10 (along with others) provides empirical context that higher-skill jobs see greater LLM exposure. Together, these supporting details give a fuller picture of how the exposure index was built and why we can trust it as a meaningful gauge of GPTs’ labor market reach.

A.2 Additional Descriptive Results

Table A.1: Summary Statistics from the American Community Survey (ACS)

	Pooled		Artist		Non-Artist	
	(1)		(2)		(3)	
	mean	sd	mean	sd	mean	sd
Annual Earnings	63271	60567	75303	75033	63223	60498
Hourly Wage	30.5	30.8	39.0	48.5	30.4	30.7
Male	0.54	0.50	0.62	0.49	0.54	0.50
Married	0.59	0.49	0.51	0.50	0.59	0.49
Age	43.5	11.3	41.5	11.1	43.6	11.3
White	0.67	0.47	0.73	0.44	0.67	0.47
Black	0.11	0.32	0.08	0.28	0.11	0.32
Asian	0.07	0.25	0.05	0.23	0.07	0.25
Years of School	14.1	3.0	15.2	2.3	14.1	3.0
Has College Degree	0.41	0.49	0.64	0.48	0.41	0.49
Observations	7408235		40208		7368027	

Notes.—Sources: American Community Survey (2017-2024). The table reports the means and standard deviations for relevant demographic and labor income variables. Earnings are deflated using the 2017 personal consumption expenditure (PCE) index.

Table A.2: LLM Exposure Index Across Artist Occupations

SOC code	Occupation	Exposure index
27-2041	Music Directors and Composers	1.694
27-1014	Multimedia Artists and Animators	0.930
27-2091	Disc Jockeys, Except Radio	0.886
27-1011	Art Directors	0.764
27-2012	Producers and Directors	0.597
27-2042	Musicians and Singers	0.454
27-1013	Fine Artists, Including Painters, Sculptors, and Illustrators	0.328
27-2032	Choreographers	-0.270
27-1012	Craft Artists	-0.321
27-2011	Actors	-0.705
27-2031	Dancers	-1.395

Table A.3: Summary Statistics on the Gallup Workplace Panel

	Pooled		Arts (ind)		Arts (occ)		Biz/Fin (occ)	
	(1)		(2)		(3)		(4)	
	mean	sd	mean	sd	mean	sd	mean	sd
AI Use - Often	0.19	0.40	0.17	0.38	0.24	0.43	0.42	0.49
AI Use - Sometimes	0.18	0.38	0.20	0.40	0.27	0.45	0.25	0.43
AI Use - Never	0.60	0.49	0.61	0.49	0.48	0.50	0.32	0.46
Hours Worked/Week	38.8	11.8	34.9	13.8	35.6	12.7	40.6	8.1
Age	43.53	14.02	43.92	14.73	41.10	13.64	41.28	12.45
Male	0.53	0.50	0.53	0.50	0.53	0.50	0.80	0.40
White	0.66	0.47	0.71	0.45	0.71	0.45	0.71	0.45
Black	0.12	0.32	0.06	0.24	0.07	0.25	0.08	0.27
Full-time	0.81	0.39	0.66	0.47	0.74	0.44	0.95	0.22
Part-time	0.19	0.39	0.34	0.47	0.26	0.44	0.05	0.22
Has Children	0.34	0.47	0.24	0.43	0.27	0.44	0.33	0.47
Married	0.53	0.50	0.46	0.50	0.47	0.50	0.57	0.49
Less than HS	0.30	0.46	0.21	0.41	0.14	0.35	0.15	0.36
Some College	0.27	0.44	0.25	0.43	0.26	0.44	0.25	0.43
College or More	0.43	0.49	0.54	0.50	0.60	0.49	0.60	0.49
Observations	70877		1909		1797		5107	

Notes.—Sources: Gallup Panel, 2023-2025. The table reports summary statistics on the key variables for the pooled sample, as well as just artists as defined by industry (Arts/Design/Entertainment/Sports/Media) and occupation (SOC = 27, Arts, Design, Entertainment, Sports, and Media Occupations) and computer and business occupations (SOC = 15) as a reference point. Respondents are given a definition of AI: “Artificial intelligence is when a computer does or enhances a task that is normally performed by a person.” AI Use - Often is based on using AI daily or a few times a week; AI Use - Sometimes is based on using AI a few times a month or a few times a year; AI Use - Never is based on using AI once a year, less than once a year, or never. Demographics include: age, race (White, Black), full-time status, having children, and being married. The sample is restricted to full and part time workers, and the observations are weighted.

Table A.4: Gallup Workplace Panel Sample Counts by Year and Artist Tag

Year	Person-wave obs.	Unique respondents	artind obs.	artocc obs.	artind unique	artocc unique
2023	9,322	9,322	284	232	284	232
2024	9,966	9,966	293	263	293	263
2025	51,589	28,922	1,332	1,302	887	822
Pooled	70,877	34,544	1,909	1,797	1,464	1,317

Notes: “Person-wave obs.” counts respondent-by-wave observations; “Unique respondents” counts distinct individuals (unique `pid`) within year. Artist tags are constructed using (i) an industry-based approach (**artind**) and (ii) an occupation-based approach (**artocc**). The pooled artist observation counts (1,909 for **artind** and 1,797 for **artocc**) correspond to the appendix summary statistics in the working paper.

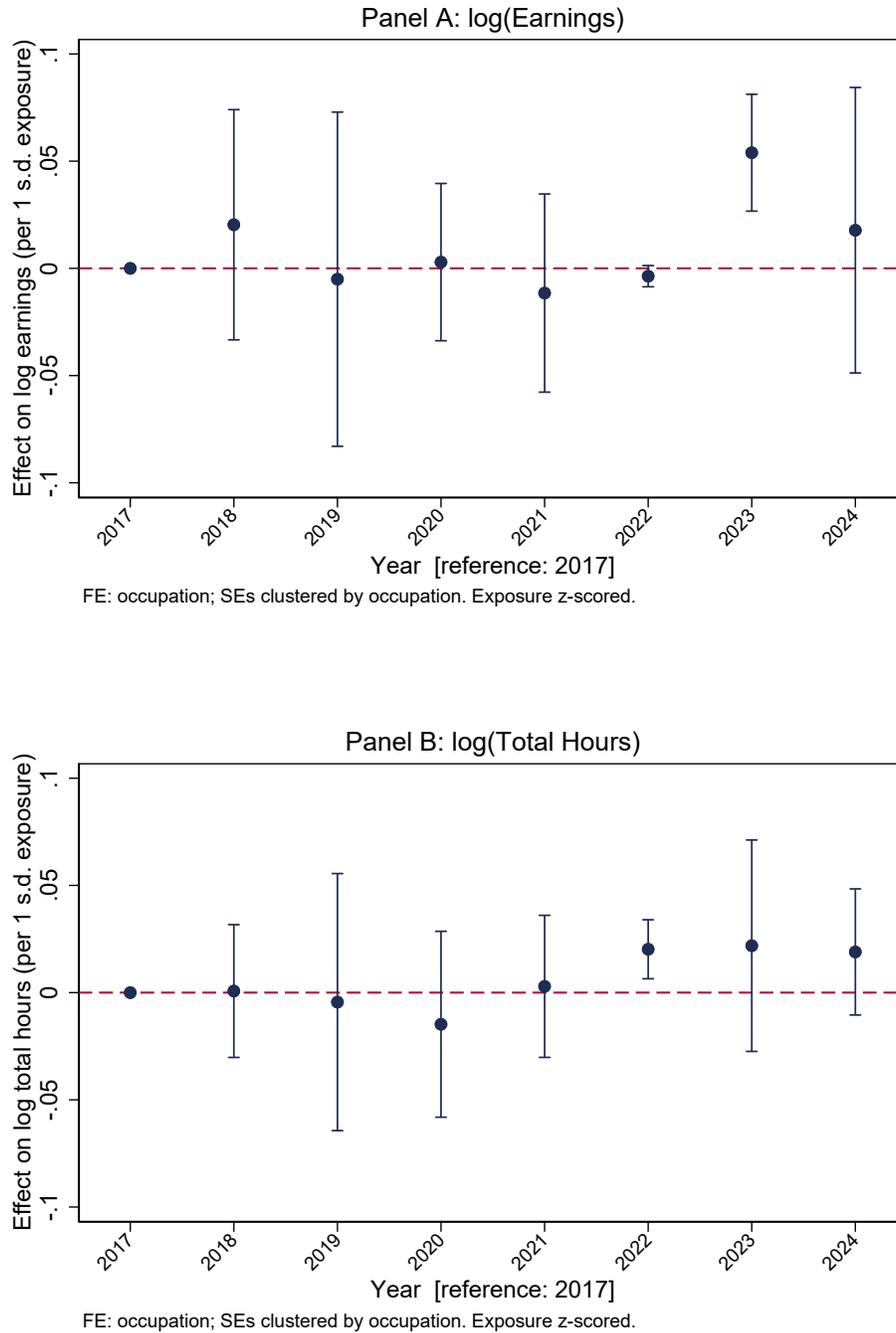
A.3 Robustness to the Main Results

Table A.5: Effects of Large Language Model Exposure on Individual Earnings

	log(Annual Earnings)		log(Annual Hours)	
	(1)	(2)	(3)	(4)
Exposure \times Year = 2017	-.051 [.044]	-.008 [.038]	.046 [.046]	.042 [.044]
Exposure \times Year = 2018	.015 [.014]	.050* [.020]	.047 [.050]	.044 [.042]
Exposure \times Year = 2019	-.052 [.046]	-.023 [.044]	.043 [.052]	.029 [.048]
Exposure \times Year = 2021	-.079* [.031]	-.041 [.031]	.060** [.014]	.050** [.014]
Exposure \times Year = 2022	-.046 [.033]	-.019 [.036]	.110** [.037]	.099* [.040]
Exposure \times Year = 2023	.119*** [.025]	.145*** [.025]	.113** [.036]	.104** [.031]
Exposure \times Year = 2024	.016 [.017]	.042 [.033]	.111** [.025]	.096** [.023]
Male		.139* [.053]		.111** [.038]
Married		.115*** [.008]		.006 [.012]
Age		.084*** [.008]		.016*** [.003]
Age squared		-.001*** [.000]		-.000*** [.000]
White		.002 [.034]		.040* [.018]
Black		-.049 [.033]		.026 [.017]
Education		.011 [.017]		.011 [.015]
Education squared		.001** [.000]		.000 [.001]
R-squared	.05	.14	.04	.06
Sample Size	28347	28347	40208	40208
Occupation FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Demographics	No	Yes	No	Yes

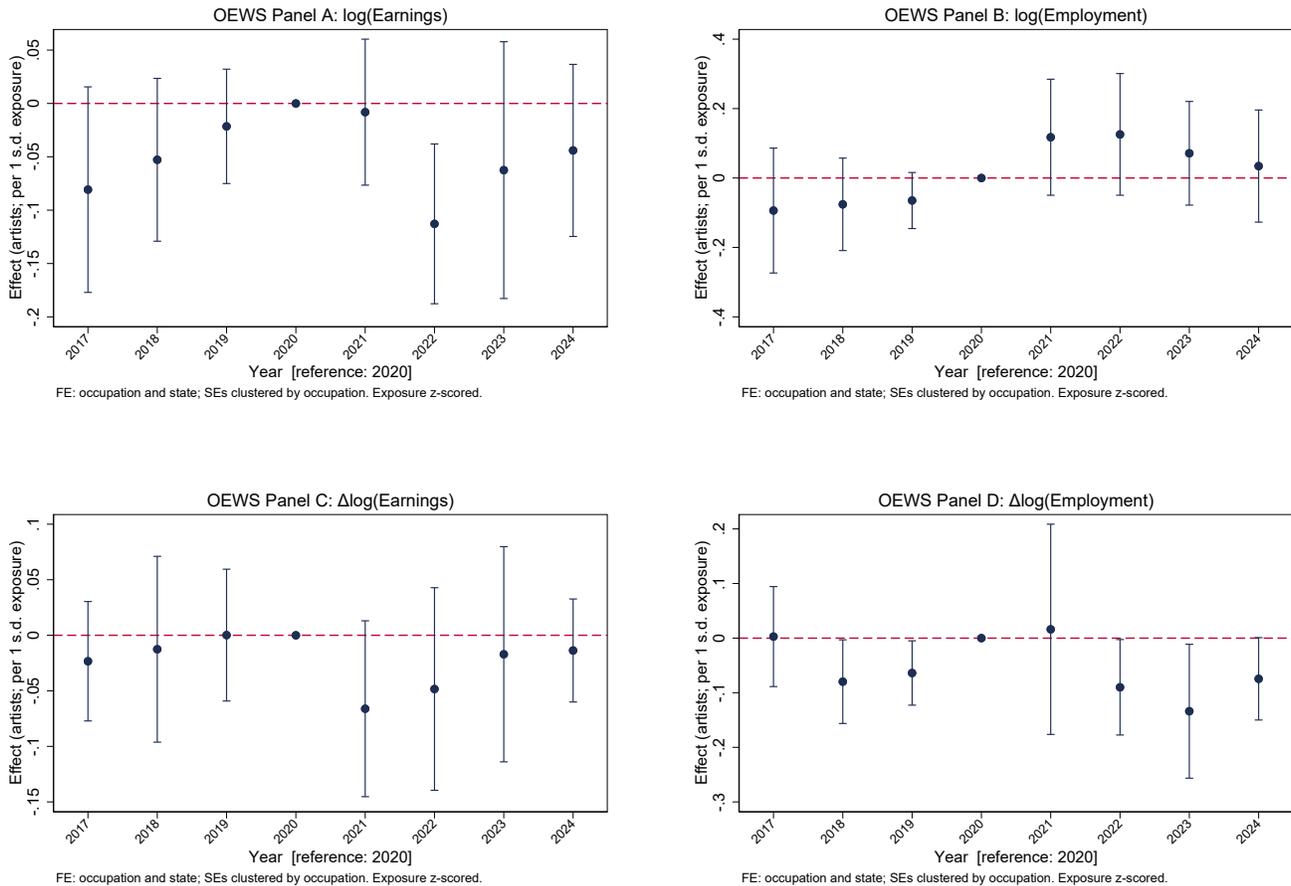
Notes.—Sources: American Community Survey (2017-2024) and [Eloundou et al. \(2024\)](#). The table reports the coefficients associated with regressions of log annual earnings (with earnings deflated with the 2017 personal consumption expenditure index) on interactions of year fixed effects and large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, conditional on state, occupation, and year fixed effects (normalized to 2020). From [Eloundou et al. \(2024\)](#), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the LLM-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Standard errors are clustered at the occupation-level.

Figure A.1: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked (Robustness Using 2017 Base Year)



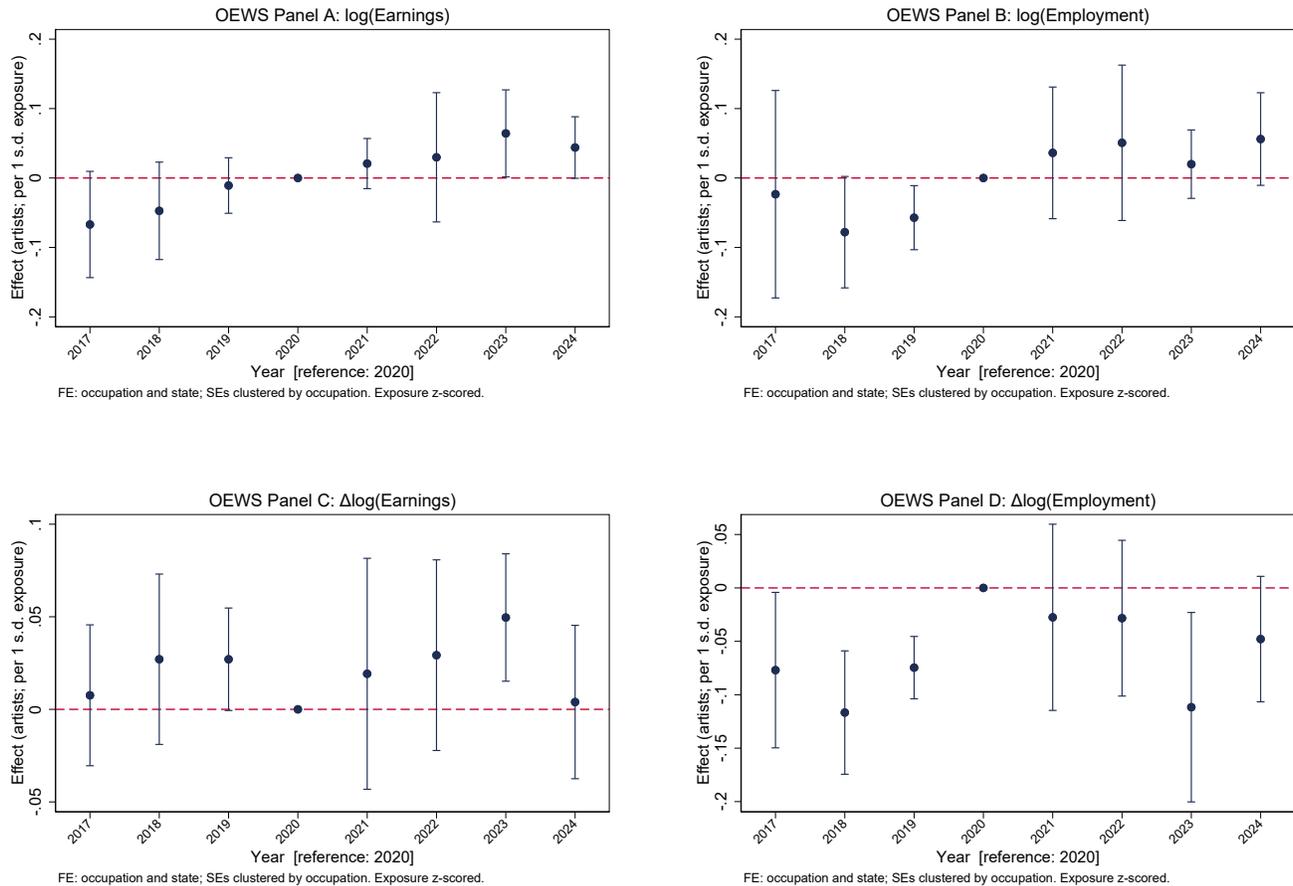
Notes.—Sources: [Eloundou et al. \(2024\)](#), American Community Survey (2017-2024). The figure reports the coefficients associated with regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z -score of β displacement risk of at least 10% of tasks (LLM-coded) from [Eloundou et al. \(2024\)](#), interacted with year dummies (normalized to 2017), conditional on SOC and year fixed effects. Demographic factors are included as controls: male, marital status, a quadratic in age, race (White, Black), and a quadratic in years of schooling. Standard errors are clustered at the occupation-level and observations are weighted by the number of employees in a state-industry cell.

Figure A.2: Dynamic Effect of Generative AI Exposure on Labor Market Outcomes (Robustness Using Human-coded Index)



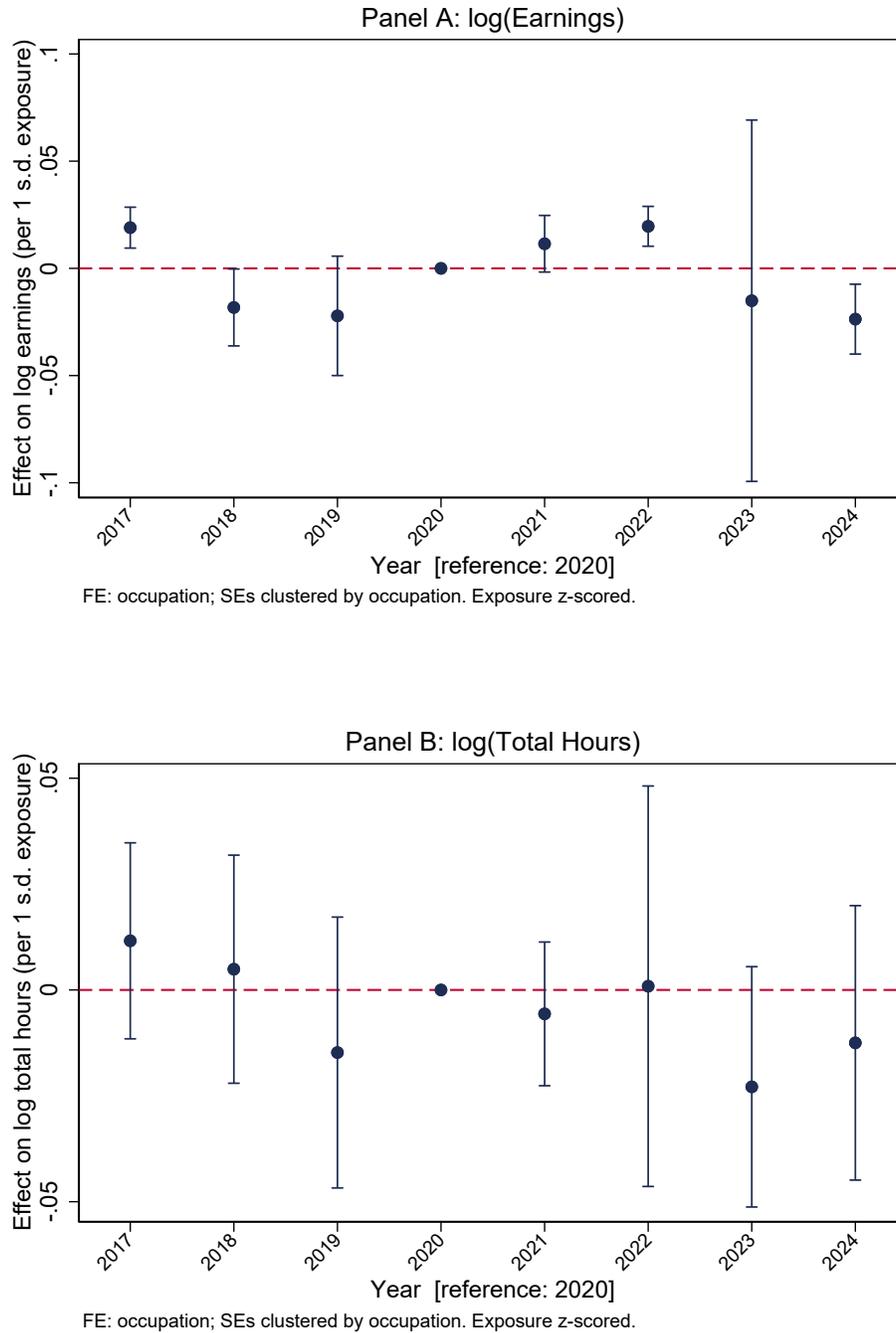
Notes.—Sources: Occupational Employment and Wage Statistics (2017-2024) and [Eloundou et al. \(2024\)](#). The figure reports the coefficients associated with regressions of year-to-year earnings and employment growth (with earnings deflated with the 2017 personal consumption expenditure index) on interactions of year fixed effects and a z-score of large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, conditional on SOC and year fixed effects (normalized to 2020). From [Eloundou et al. \(2024\)](#), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the human-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Employment and earnings growth is winsorized at the top and bottom percentile. Standard errors are clustered at the occupational-level.

Figure A.3: Dynamic Effect of Generative AI Exposure on Labor Market Outcomes (Robustness Using Artist \times Exposure)



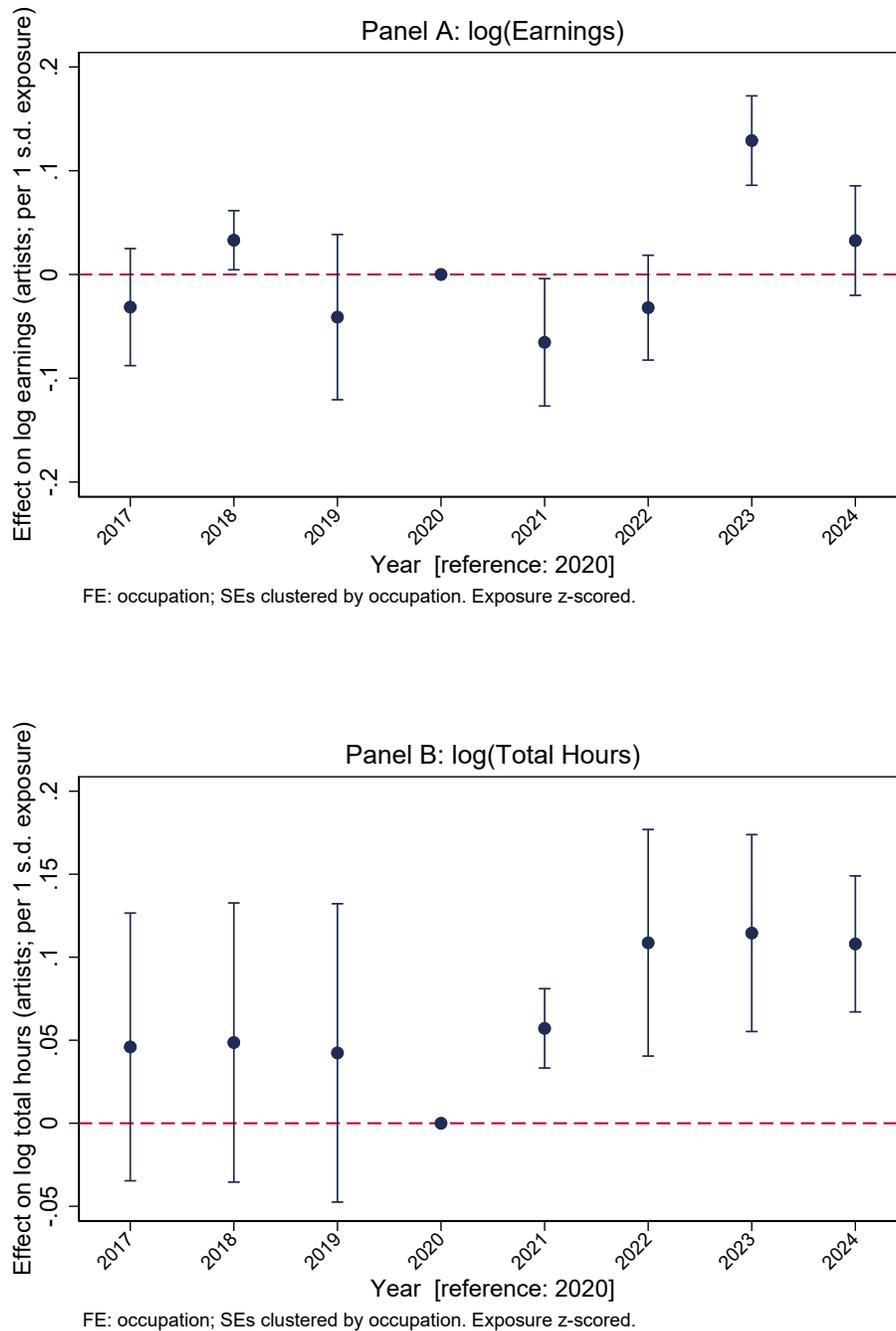
Notes.—Sources: Occupational Employment and Wage Statistics (2017-2024) and [Eloundou et al. \(2024\)](#). The figure reports the coefficients associated with regressions of year-to-year earnings and employment growth (with earnings deflated with the 2017 personal consumption expenditure index) on interactions of year fixed effects, an indicator with being an artist, and a z -score of large language model (LLM) occupational exposure at the six-digit standard occupational classification (SOC) level, conditional on SOC and year fixed effects (normalized to 2020). From [Eloundou et al. \(2024\)](#), experts assessed each occupational task on a defined rubric, rating the extent to which LLMs such as GPT models could feasibly execute or significantly assist with the task (using the LLM-coded version). This expert judgment involved assessing the compatibility of each task with current generative AI capabilities, resulting in an occupation-level exposure score. The parameter β that is used to measure exposure is interpreted as the percentage of job tasks within each occupation that LLMs could feasibly perform, augment, or significantly assist with minimal human intervention. The sample is restricted to artists, defined as: (a) art directors, craft artists, fine artists, special effects artists, and related (SOC 27-101), (b) actors, producers, directors (SOC 27-201), (c) dancers and choreographers (SOC 27-203), (d) musicians, singers, and related (SOC 27-204), and (e) entertainers, performers, and related (SOC 27-209). Employment and earnings growth is winsorized at the top and bottom percentile. Standard errors are clustered at the occupational-level.

Figure A.4: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked (Robustness with Human-coded Exposure)



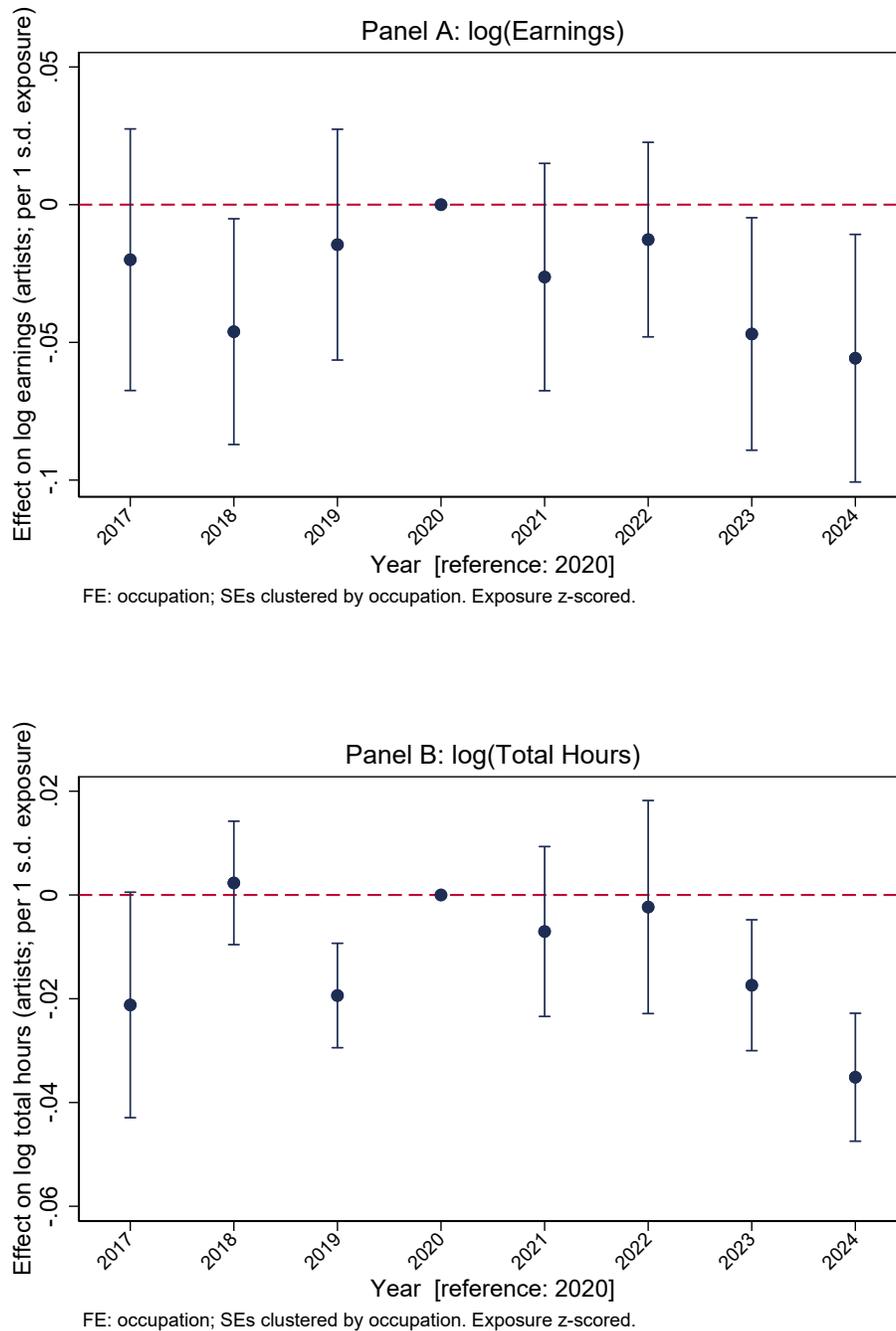
Notes.—Sources: ?, American Community Survey (2017-2024). The figure reports the coefficients associated with regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z -score of β displacement risk of at least 10% of tasks (human-coded) from ?, interacted with year dummies (normalized to 2020), conditional on SOC and year fixed effects. Demographic factors are included as controls: male, marital status, a quadratic in age, race (White, Black), and a quadratic in years of schooling. Standard errors are clustered at the occupation-level and observations are weighted by the number of employees in a state-industry cell.

Figure A.5: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked (Robustness with Artist \times Exposure Interaction)



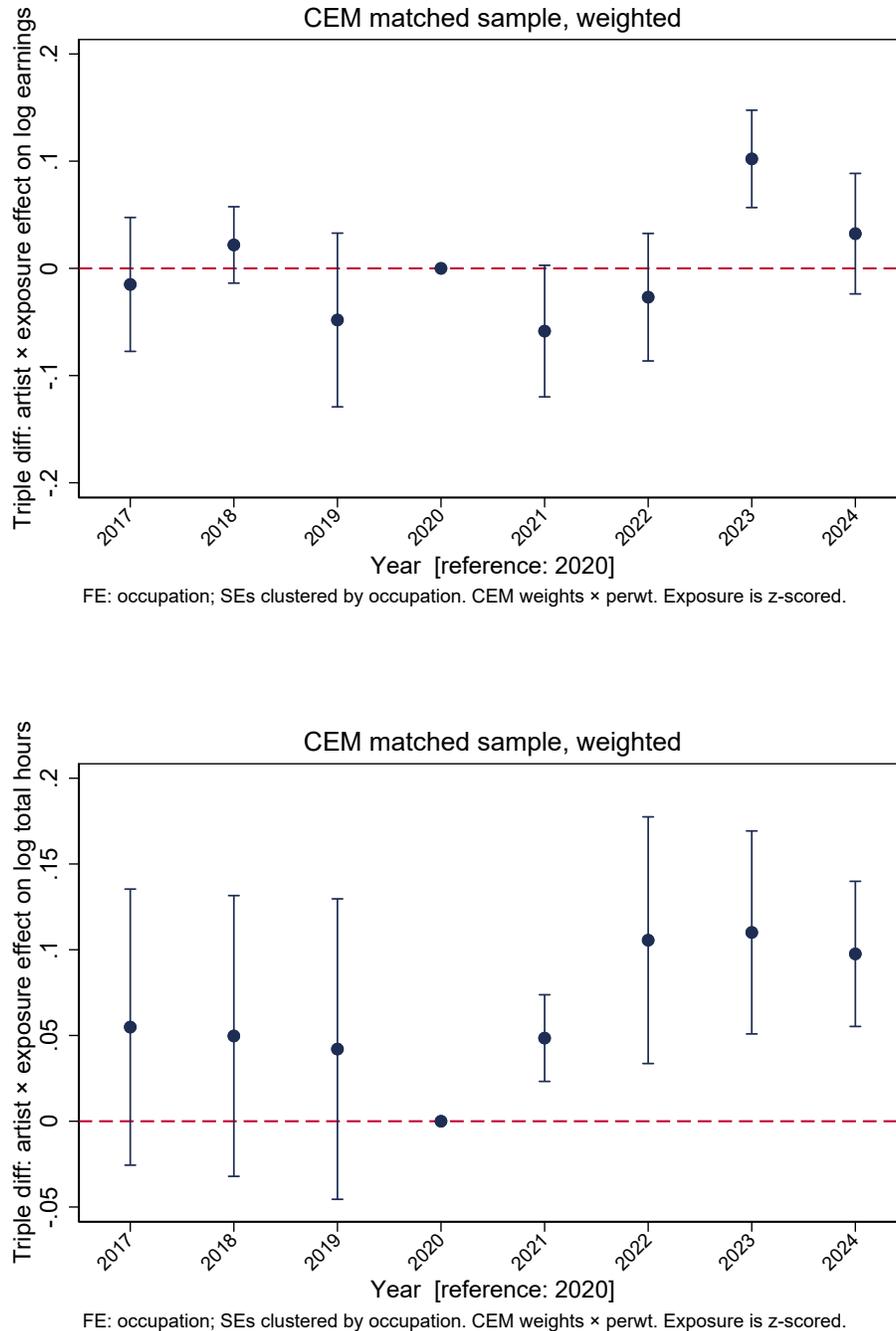
Notes.—Sources: ?, American Community Survey (2017-2024). The figure reports the coefficients associated with regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z -score of β displacement risk of at least 10% of tasks (LLM-coded) from ?, interacted with year dummies (normalized to 2020) and an indicator for being an artist, conditional on SOC and year fixed effects. Demographic factors are included as controls: male, marital status, a quadratic in age, race (White, Black), and a quadratic in years of schooling. Standard errors are clustered at the occupation-level and observations are weighted by the number of employees in a state-industry cell.

Figure A.6: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked (Robustness Using Felten et al. (2021) with Artist \times Exposure Interaction)



Notes.—Sources: Felten et al. (2021), American Community Survey (2017-2024). The figure reports the coefficients associated with regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z-score of Felten et al. (2021) at the county-level, interacted with year dummies (normalized to 2020) and an indicator for being an artist, conditional on SOC, county, and year fixed effects. Demographic factors are included as controls: male, marital status, a quadratic in age, race (White, Black), and a quadratic in years of schooling. Standard errors are clustered at the occupation-level and observations are weighted by the number of employees in a state-industry cell.

Figure A.7: Dynamic Effect of Generative AI Exposure on Worker Earnings and Hours Worked (Robustness Using Coarsened Exact Matching)



Notes.—Sources: [Eloundou et al. \(2024\)](#), American Community Survey (2017–2024). The figure reports coefficients from event study regressions of log annual earnings (deflated using the 2017 personal consumption expenditure index) and log total hours worked on a z -score of β displacement risk (LLM-coded) from [Eloundou et al. \(2024\)](#), interacted with year indicators (normalized to 2020) and an indicator for being an artist. The sample is reweighted using coarsened exact matching (CEM) matching artists with others within the broader two-digit SOC group 27 to improve comparability. Matching is conducted exactly on age (five-year bins from 25 to 65) and education (high school or less, some college or associate degree, and college or more), and matched observations are weighted by the product of ACS person weights and CEM weights. All specifications include occupation fixed effects, demographic controls (sex, marital status, race, and quadratic terms in age and years of schooling), and standard errors clustered at the occupation level.