WORKING PAPER

Estimating the Task Content of Work: Workforce Design for AI-Driven Human–Robot Collaboration in Intralogistics

Pierre Bouquet[a], Nicolo Bagnoli[a], Yossi Sheffi[a]

[a]Massachusetts Institute of Technology (MIT), Center for Transportation & Logistics, 1 Amherst Street, Cambridge, MA, USA;

ARTICLE HISTORY
Compiled December 3, 2025

ABSTRACT

This paper addresses the challenge of strategic workforce planning for AI-driven human–robot collaboration (AI-HRC) in intralogistics. We ask two questions: how can task-level full-time equivalent (FTE) estimates be constructed from existing labor statistics, and how can these estimates, combined with AI exposure metrics, inform strategic AI-HRC design and workforce planning? Drawing on U.S. Bureau of Labor Statistics employment data, O*NET occupational profiles, and task-level AI exposure scores, we develop a stochastic task–time framework that decomposes occupations into tasks and models task frequencies as probability vectors on the simplex. A covariance-completion procedure reconstructs task covariance matrices consistent with survey standard errors, enabling the translation of occupational data into task-level and detailed work activity (DWA)-level FTE estimates with uncertainty bounds. Applying the framework to the U.S. intralogistics workforce, we find that roughly 380,000 FTEs (about 17% of workers) are concentrated in DWAs in the top 15% of AI-exposed DWAs. These results provide task-specific insight into AI-driven automation and support scenario-based workforce planning by linking alternative AI-HRC adoption paths to task-level FTE impacts, uncertainty bands, and upskilling priorities, thereby offering an analytical foundation for resilient, human-centered AI-HRC systems.

KEYWORDS
Human-robot collaboration; Intralogistics; Task-based workforce modeling; Task-based automation; Uncertainty quantification; Workforce design

Pierre Bouquet. Email: pibou149@mit.edu

1. Introduction

Artificial intelligence (AI) is reshaping work organization, producing heterogeneous effects across tasks, occupations, and industries. Recent audits indicate that approximately 80 percent of the U.S. workforce has at least 10 percent of tasks exposed to LLM capabilities or LLM-powered software, while about 19 percent has 50 percent or more of tasks exposed (Eloundou et al. 2024). Beyond exposure studies, empirical evidence from thousands of real interactions demonstrates high applicability for information-gathering, writing, and advising tasks, particularly where communication and coordination are essential (Tomlinson et al. 2025). Task-based economic research finds that digital technologies often replace routine tasks but augment non-routine problem solving (Autor, Levy, and Murnane 2003; Autor and Dorn 2013). These findings underline the importance of task-level analysis of AI-driven automation. To further understand the impact of AI, it is necessary to embed it within physical task systems, integrating sensing, actuation, and human judgment to enable coordination between robots and frontline workers. This integration is especially relevant in the operational context of logistics.

The logistics sector is undergoing a significant transformation as AI-driven technologies reshape transportation, warehousing, and distribution. Performance in logistics depends on synchronous, high-volume task execution and must adapt to volatility from demand peaks, SKU proliferation, and service-level pressures. Warehousing has shifted from picker-to-parts systems to robotized fulfillment centers, resulting in higher service quality and reduced lead times (Boysen and De Koster 2025). Industry 4.0 technologies and digital twins reduce information lags and enable tighter synchronization between production and warehousing (Li et al. 2023; De Koster et al. 2025). Digital twins are best understood as cyber-physical systems that maintain real-time, data-driven representations of logistics processes through continuous streams of IoT and operational data (Park, Son, and Noh 2021). By 2028, approximately 80 percent of warehouses and distribution centers are projected to use some form of robotics or automation, highlighting the rapid pace and scale of change (Ma and Saénz 2025). Autonomous mobile robots (AMRs) and robotic mobile fulfillment systems (RMFS) are seeing growing adoption in intralogistics, as more facilities integrate them into their planning and control processes (Fragapane et al. 2021). However, the sector continues to rely heavily on human workers, particularly in picking, sorting, and exception handling, where variability, damage, and unique items constrain the potential for full automation (U.S. Bureau of Labor Statistics 2024).

These technological shifts have set the stage for AI-driven human-robot collaboration (AI-HRC) within logistics facility operations. This sector - commonly referred to as intralogistics - concerns the management of material and information flows in warehouses, distribution centers, and manufacturing plants. Cobots, AMRs, and perception systems bring speed and precision, while humans provide dexterity, contextual understanding, and exception management (Pasparakis, de Vries, and de Koster 2023). When well-designed, such collaboration enhances job satisfaction and self-efficacy; when poorly paced or interfaced, it can erode trust and underutilize both humans and robots (Pasparakis, de Vries, and de Koster 2023). Despite rapid automation, humans remain the backbone of system resilience. In this evolving context, AI-enabled collaboration does not simply replace labor but reshapes it.

To address the evolving nature of work in intralogistics and the challenges posed by AI-HRC and resilience requirements, this study proposes a stochastic task-time framework for strategic workforce planning. The framework generates estimates of the total number of FTEs performing each task, including confidence levels. These estimates are then integrated with AI automation exposure studies to inform AI-HRC workforce planning. Methodologically, this approach extends prior work by Cai et al. (2025) by defining the intralogistics industry as a set of job-task pairs and incorporating uncertainty bounds into functionals of the frequency vector developed by Martin and Monahan (2022).

The results are estimated ranges of the number of FTEs assigned for each task within the U.S. intralogistics workforce. Although the primary focus is on the U.S. intralogistics sector, the methodology is broadly applicable and can be replicated for any workforce. When combined with AI exposure studies, this work can support strategic workforce planning by identifying the human capital required for AI-HRC tasks. To enhance practical relevance, exposures are interpreted in relation to AMR/RMFS, IoT instrumentation, and digital-twin design patterns. We also provide a public repository with all the code and data pipelines, which can be adapted to any workforce representation.

The remainder of the paper is organized as follows. Section 2 reviews the literature and positions this work among related studies. Section 3 describes the data and methodologies developed for the stochastic task-time framework. Section 4 presents the results and the diagnostics of the methodology. Section 5 examines the implications for strategic workforce planning and AI-HRC design, and discusses the assumptions underlying this study. Section 6 summarizes key insights and outlines directions for future research.

2. Literature Review

We present a comprehensive literature review on AI-driven automation, focusing on intralogistics. The section is organized as follows: first, modeling technological impact with task-based workforce models; next, presenting empirical evidence on AI innovations; finally, reviewing current AI-driven technologies in intralogistics and AI-HRC.

2.1. *Modeling the Impact of Technological Innovation Through Task-Based Workforce Models*

The relationship between technological progress and the reorganization of work has been studied for decades using task-based economics. Foundational work formalized how computers substitute for routine tasks and complement non-routine cognitive and social functions, giving rise to labor-market polarization analyses (Autor, Levy, and Murnane 2003; Autor and Dorn 2013). Subsequent work extended this framework to automation and AI, showing that new technologies both displace and create task frontiers (Autor 2015; Acemoglu and Restrepo 2019; Autor et al. 2024). This decomposition of occupations into constituent tasks, and the evaluation of their susceptibility to technological change, has become standard in labor economics and management studies (Frey and Osborne 2017; Brynjolfsson and Mitchell 2017; Felten, Raj, and Seamans 2018; Eloundou et al. 2024).

2.2. *Empirical Evidence on the Effects of AI Innovations*

A parallel body of empirical literature quantifies the effects of AI-driven innovations on productivity and work quality. Randomized and quasi-experimental studies across multiple domains demonstrate substantial but heterogeneous gains that depend on task structure and worker experience. In software engineering, controlled trials of code-assistance tools show measurable reductions in completion time and error rates (Peng et al. 2023). In text-based knowledge work, LLM-based writing assistants boost speed and quality while narrowing performance dispersion among workers (Noy and Zhang 2023). Field studies in customer support and consulting suggest that generative AI improves both throughput and consistency, particularly among less experienced employees (Brynjolfsson, Li, and Raymond 2025; Dell'Acqua et al. 2023). Collectively, these studies indicate that AI-driven technologies are primarily redistributive rather than purely additive. They elevate lower performers, reduce productivity variance, and shift expertise toward supervision and integration with existing workflows, rather than completely redefining them. These behavioral mechanisms influence how AI exposure translates into performance gains and establish a baseline for understanding the augmentation enabled by this technology.

2.3. *AI-driven Technologies in Intralogistics and HRC*

While most AI–work studies focus on digital services, intralogistics offers a crucial testing ground where information processing and physical execution converge. Across successive automation waves, warehouse systems have evolved from human-supervised information platforms to nearly autonomous decision-making layers, computerizing functions such as navigation, scheduling, routing, and replenishment (Boysen and De Koster 2025). AMRs now coordinate movements once assigned manually, RMFS manage retrieval tasks formerly handled by forklift or pallet operators, and IoT-based predictive maintenance automates monitoring and fault prediction previously performed by technicians (Li et al. 2023). The digital and physical layers have effectively merged - each pick, scan, and movement generates data that feeds learning algorithms for continuous optimization. Industry 4.0 instrumentation and digital-twin synchronization reduce information latency and tightly couple production with intralogistics in near real time, supporting scenario analysis, capacity planning, and resilience testing under disruptions (Li et al. 2023; De Koster et al. 2025; Liu, Pan, and Ballot 2024).

Within this technological transition, AI-HRC has become the organizing principle of next-generation intralogistics systems. Research based on current technology increasingly supports the superiority of "humans-in-the-loop" configurations over full automation in high-variability contexts, as anomalies and changeovers remain difficult to codify (De Koster and Roy 2024; De Koster et al. 2025). In this context, robots perform high-speed, repetitive, and physically demanding tasks, while humans contribute dexterity, contextual awareness, and adaptive decision-making (Perotti et al. 2025; Pasparakis, de Vries, and de Koster 2023). As tasks are divided in AI-HRC, even minor adjustments in automation design can influence workforce composition (Autor and Thompson 2025). Traditional workforce metrics at the occupational level obscure the inherent variability and uncertainty present in mixed human–AI task systems (Nourmohammadi et al. 2025).

3. Methodology

This section introduces the workforce modeling framework, followed by a description of the data sources and preprocessing steps. We outline the statistical approach for estimating task times and associated uncertainties. Finally, these estimates are translated into FTEs and aggregated to inform AI-HRC strategic workforce planning.

3.1. *Overview and Framework*

The framework developed in this study provides a unified set of processes for translating occupational-level data into task-level estimates for AI-HRC workforce planning. The framework consists of five structured stages, as illustrated in Figure 1.

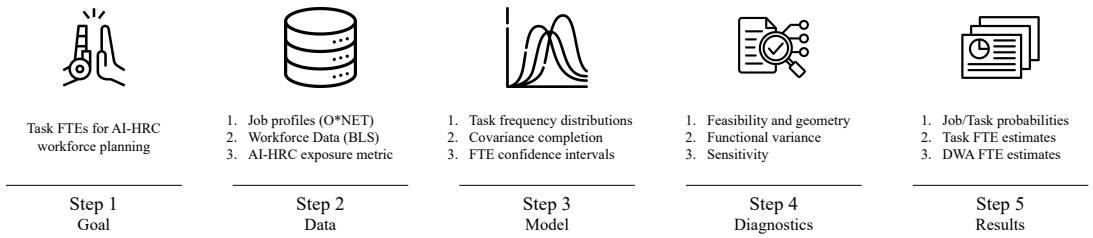

Figure 1.: A modeling framework for estimating the number of workers per task within the intralogistics sector.

The model integrates three key data sources: (i) intralogistics job employment data; (ii) detailed task-job relationships; and (iii) metrics of task exposure to AI. The model translates occupational data into task-level FTE ranges. Diagnostic tests validate the model's internal consistency. The outputs are interpreted along three dimensions: aggregation of FTEs at the task level, integration of AI exposure scores to identify tasks most suitable for AI-HRC, and strategic workforce planning. The methodological assumptions underlying this framework are stated throughout Section 3 and summarized in Appendix A. A detailed discussion of limitations, scope conditions, and sources of uncertainty is provided in Appendix B.

3.2. *Data Sources*

This section introduces the study data, specifying their sources, structure, and preprocessing steps.

3.2.1. BLS Employment and Wage Data

We obtained employment data for the Transportation and Warehousing sector from the Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) program. The OEWS is a semiannual survey covering approximately 830 occupations based on a sample of approximately 1.1 million establishments representing 55 percent of total U.S. employment (U.S. Bureau of Labor Statistics 2024). The OEWS reports employment counts by occupation along with standard errors derived from the survey design. This paper employs the May 2024 OEWS release.

*3.2.2. O*NET Occupational Data*

The O*NET serves as the leading repository of standardized occupational data in the United States and a foundation for workforce planning strategies. Launched in 1998, it is maintained by the U.S. Department of Labor's Employment and Training Administration. The O*NET Content Model organizes occupational information hierarchically (National Center for O*NET Development 2024). The database includes about 55,000 job titles which are consolidated into approximately 900 detailed occupational profiles. Each profile breaks an occupation into work elements organized hierarchically across several precision levels. At the finest granularity, over 18,000 unique task statements describe specific activities workers perform. They are grouped into about 2,000 Detailed Work Activities (DWAs). The structure of the data, survey protocols, and extracts of the data tables used are provided in Appendix C. This study utilizes O*NET version 28.3, released in May 2024.

3.3. *AI Exposure Metrics by Task*

We adopt the AI exposure metrics proposed by Bouquet, Kaboli, and Sheffi (2025). This methodology embeds DWA descriptions using the SBERT model and compares them against a large corpus of news articles referencing automation, robotics, and artificial intelligence. While cosine similarity determines the semantic relevance of an article to a task, the methodology employs a "factual relevance" filter derived from the RoBERTa sentiment model to distinguish actual technological capability from media speculation. The metric explicitly prioritizes neutral sentiment: in the context of technological news, neutrality serves as a proxy for objective reporting on actual deployments (e.g., "The warehouse implemented AMRs"), whereas polarized sentiment typically reflects editorializing (e.g., "Robots will save/destroy the workforce"). By isolating the signal of technical feasibility from the noise of media speculation, the model calculates a cosine-weighted average of these objectivity-filtered scores over a two-year rolling window. This approach ensures comparability across DWAs and robustness to short-term fluctuations in news coverage. The result is a continuous score between 0 and 1 for each DWA, indicating the potential exposure to AI-driven automation.

3.4. *Data Preparation*

The process begins with the O*NET task ratings file. A sequence of cleaning steps is then applied to prepare the data for modeling and to ensure that all valid and available information is used.

(1) **Preliminary cleaning** involved dropping rows flagged by O*NET as "Recommend Suppress = Y" and removing records with missing data values.
(2) **Category completeness** was enforced by requiring complete data for each task in the dataset. Tasks with missing data were excluded from further analysis.

3.5. *Statistical Modeling*

This section introduces a framework for estimating the number of FTEs performing tasks within an organization, sector, or economy under uncertainty.

3.5.1. Task Frequency Distributions

Data on task execution frequency is derived from the O*NET data collection program, which administers structured surveys to job incumbents using a two-stage response protocol. First, respondents indicate whether a specific task is relevant to their occupation using a binary filter. If the task is deemed relevant, respondents then report the frequency of performance using a seven-point ordinal scale, ranging from "Yearly or less" to "Hourly or more." O*NET aggregates these responses to report the percentage of workers selecting each category, along with design-based standard errors. Further details on the data structure and survey protocols are provided in Appendix C.2.

Let i index occupations and k tasks. Denote by $\widehat{RT}_{i,k}$ the design-based percentage of incumbents who mark task k as relevant in occupation i, with standard error $\mathrm{SE}(\widehat{RT}_{i,k})$. We convert to probabilities via

$$\hat{R}_{i,k} = \frac{\widehat{RT}_{i,k}}{100}, \qquad \mathrm{Var}(\hat{R}_{i,k}) = \left(\frac{\mathrm{SE}(\widehat{RT}_{i,k})}{100}\right)^2 \tag{1}$$

Similar to Martin and Monahan (2022), we introduce an explicit unconditional "never" category (indexed $r = 0$) as the complement of relevance,

$$\hat{p}_{i,k,0} = 1 - \hat{R}_{i,k}, \qquad \mathrm{Var}(\hat{p}_{i,k,0}) = \mathrm{Var}(\hat{R}_{i,k}). \tag{2}$$

Conditioned on relevance, O*NET reports the design-based estimator of the percentage of workers choosing each frequency category $r = 1, \ldots, 7$, denoted $\widehat{FT}_{i,k,r}$ with standard errors $\mathrm{SE}(\widehat{FT}_{i,k,r})$. We first rescale them:

$$\hat{F}_{i,k,r} = \frac{\widehat{FT}_{i,k,r}}{100}, \qquad \mathrm{Var}(\hat{F}_{i,k,r}) = \left(\frac{\mathrm{SE}(\widehat{FT}_{i,k,r})}{100}\right)^2, \qquad \sum_{r=1}^{7} \hat{F}_{i,k,r} = 1. \tag{3}$$

Combining relevance with conditional frequencies yields the unconditional category probabilities for a randomly selected worker:

$$\hat{p}_{i,k,r} = \begin{cases} 1 - \hat{R}_{i,k}, & r = 0, \\ \hat{R}_{i,k}\,\hat{F}_{i,k,r}, & r = 1, \ldots, 7, \end{cases} \qquad \sum_{r=0}^{7} \hat{p}_{i,k,r} = 1. \tag{4}$$

Assuming the design-based estimators $\hat{R}_{i,k}$ and $\hat{F}_{i,k,r}$ are statistically independent (elicited in distinct survey steps), we compute the exact variance of $\hat{p}_{i,k,r}$ using the product variance formula from Goodman (1960). For $r \geq 1$, this yields:

$$\mathrm{Var}(\hat{p}_{i,k,r}) \;=\; \hat{F}_{i,k,r}^2\,\mathrm{Var}(\hat{R}_{i,k}) \;+\; \hat{R}_{i,k}^2\,\mathrm{Var}(\hat{F}_{i,k,r}) \;+\; \mathrm{Var}(\hat{R}_{i,k})\,\mathrm{Var}(\hat{F}_{i,k,r}). \tag{5}$$

If any $\hat{p}_{i,k,r}$ evaluates to zero after construction (or its standard error is numerically zero), we add $\epsilon = 10^{-6}$ for computational stability to the affected entries and renormalize $\hat{\mathbf{p}}_{i,k}$ to maintain the simplex constraint.

To translate the unconditional category probabilities into expected annual occurrence counts, we use an annualization vector based on the work by Tomlinson et al. (2025).

$$\mathbf{w} = \big[0,\ 1,\ 4,\ 24,\ 104,\ 260,\ 780,\ 2080\big]^\top, \tag{6}$$

where each entry corresponds to the O*NET scale [Never, Yearly, More-than-yearly, More-than-monthly, ..., Hourly-or-more]. These values implement conservative lower-end midpoints of the "more than" intervals (e.g., $4 \approx 2\times$ per half-year; $24 \approx 2\times$ per month) and standard full-time assumptions (hourly $\approx$ 2080 work-hours/year). Finally, let $\hat{\mu}_{i,k}$ be the plug-in estimator of the expected annual number of occurrences for task k in occupation i:

$$\hat{\mu}_{i,k} = \mathbf{w}^\top \hat{\mathbf{p}}_{i,k} = \sum_{r=0}^{7} w_r\, \hat{p}_{i,k,r}. \tag{7}$$

While the O*NET survey uses a complex design, we adopt a model-based approach for uncertainty analysis. We treat the empirically derived probability vector $\hat{\mathbf{p}}_{i,k}$ as the latent population distribution and assume each worker's categorical response is an independent and identically distributed draw from $\hat{\mathbf{p}}_{i,k}$.

3.5.2. Covariance Completion

Determining the uncertainty of the annualized task frequency, $\hat{\mu}_{i,k} = \mathbf{w}^\top \hat{\mathbf{p}}_{i,k}$, requires the full covariance matrix $\widehat{\Sigma}_{i,k} \in \mathbb{S}^8$, the space of symmetric 8×8 matrices.. However, O*NET survey data provides only the marginal variances, leaving the cross-category correlations undefined. To recover a valid covariance structure, we formulate a convex completion problem that enforces consistency with both the observed survey design and the geometry of the probability simplex (Grussler, Rantzer, and Giselsson 2018).

Formally, any covariance matrix of a probability vector on the unit simplex must satisfy the tangent space constraint $\Sigma\mathbf{1} = \mathbf{0}$ (Aitchison 1982). We estimate $\widehat{\Sigma}_{i,k}$ by selecting the positive semidefinite (PSD) matrix that satisfies this geometric constraint and matches the reported marginals, while minimizing the Frobenius norm $\|\Sigma\|_F$. This objective function acts as a parsimonious regularizer (Horn and Johnson 2012; Boyd and Vandenberghe 2004): it recovers the minimum-energy covariance structure required to satisfy the constraints, thereby avoiding the imposition of spurious correlations. This approach is robust to survey noise, as Frobenius norm minimization prevents the amplification of sampling errors into the off-diagonal terms (Yue et al. 2024).

All semidefinite programs are solved with MOSEK (2025); solver settings are reported in Appendix D.

Let $\mathbf{v}_{i,k} = [v_{i,k,0}, \ldots, v_{i,k,7}]^\top$, where $v_{i,k,r} = \mathrm{Var}(\hat{p}_{i,k,r})$ denotes the reported marginal variance for frequency category r of task k in occupation i. We obtain the

covariance matrix $\widehat{\Sigma}_{i,k}$ as:

$$\begin{aligned} \widehat{\Sigma}_{i,k} \;\in\; \arg\min_{\Sigma\in\mathbb{S}^8} \;& \|\Sigma\|_F, \\ \text{s.t.} \quad & \Sigma \succeq 0, \\ & \operatorname{diag}(\Sigma) = \mathbf{v}_{i,k}, \\ & \Sigma\mathbf{1} = \mathbf{0}. \end{aligned} \tag{8}$$

Here, $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbb{S}^8$ the space of symmetric 8×8 matrices, and $\Sigma\mathbf{1} = \mathbf{0}$ enforces the unit-sum property of probabilities. The variance of the plug-in estimator $\hat{\mu}_{i,k} = \mathbf{w}^\top \hat{\mathbf{p}}_{i,k}$ follows from the law of propagation of variances (Bevington and Robinson 2003)

$$\operatorname{Var}(\hat{\mu}_{i,k}) \;=\; \mathbf{w}^\top \widehat{\Sigma}_{i,k}\mathbf{w}, \tag{9}$$

3.5.3. Model Diagnostics

We assess the completed covariance matrices $\widehat{\Sigma}_{i,k}$ along three dimensions: feasibility, geometry on the simplex, and robustness.

3.5.3.1. Feasibility and Simplex Geometry. We first verify the solver convergence for each solution of problem (8), since there is a potential for ill-conditioning and numerical instability in PSD problems (Wolkowicz, Saigal, and Vandenberghe 2000). Hence we check the feasibility of each converged solution, based on two principal criteria.

The simplex tangent constraint is imposed exactly: each estimated covariance $\widehat{\Sigma}_{i,k}$ must satisfy $\widehat{\Sigma}_{i,k}\mathbf{1} = \mathbf{0}$, ensuring compatibility with compositional geometry (Aitchison 1982). Positive semidefiniteness is verified numerically by requiring $\lambda_{\min}(\widehat{\Sigma}_{i,k}) \geq -10^{-8}$, consistent with standard solver tolerances and reflecting typical floating-point precision (Wolkowicz, Saigal, and Vandenberghe 2000).

Solutions are deemed feasible only if both criteria are met; otherwise, the task is rejected from the model. For additional rigor and reference, we report supplementary diagnostics, including diagonal agreement, total uncertainty, and effective rank, in Appendix E.1.

3.5.3.2. Functional Variance and Feasible Bounds. To link our diagnostics to inference on annual occurrences, we compute the variance of the plug-in estimator, $\hat{\mu}_{i,k}$, as $\mathbf{w}^\top\widehat{\Sigma}_{i,k}\mathbf{w}$. We benchmark this value against the feasible variance interval $[\sigma^2_{\min}, \sigma^2_{\max}]$, which is obtained by solving two auxiliary problems to minimize and maximize $\mathbf{w}^\top\Sigma\mathbf{w}$, subject to $\Sigma \succeq 0$, $\operatorname{diag}(\Sigma) = \mathbf{v}_{i,k}$, and $\Sigma\mathbf{1} = 0$ (Boyd and Vandenberghe 2004). We report the normalized position

$$\frac{\mathbf{w}^\top\widehat{\Sigma}_{i,k}\mathbf{w} - \sigma^2_{\min}}{\sigma^2_{\max} - \sigma^2_{\min}} \in [0,1], \tag{10}$$

which quantifies the conservativeness of the completion relative to the range of all feasible solutions under these constraints. To prevent a small number of extreme yet feasible completions from dominating uncertainty budgets, we exclude tasks whose

normalized position lies outside the interval $[0.05, 0.95]$, following robust-trimming principles in Huber and Ronchetti (2009); see Appendix E.2 for details.

3.5.3.3. Sensitivity. Finally, we assess robustness to sampling-error misspecification by perturbing the diagonal targets $\mathbf{v}_{i,k}$ by $\pm 5\%$ (Huber and Ronchetti 2009; Saltelli et al. 2004), recomputing the completion, and recording the absolute and relative change in $\mathbf{w}^\top \widehat{\Sigma}_{i,k}\mathbf{w}$. Small changes indicate that the inferred functional uncertainty is both numerically stable and substantively robust to moderate misspecification of the marginals that could come from O*NET sampling errors (Yuan and Zhang 2013). For each task, we report the maximum and minimum changes observed, as well as the spread and symmetry of the variance estimates. Tasks exhibiting numerically unstable sensitivity, specifically, those with symmetric relative change exceeding 0.1 after a perturbation of 5% of the marginals are excluded from further analysis; see Appendix E.3 for details.

3.6. *From Tasks to Full-Time Equivalents*

The preceding sections established methods for estimating task occurrence rates $\hat{\mu}_{i,k}$ with full covariance structures. We now translate these annual occurrence estimates into workforce-level metrics. This process involves three stages. First, raw annual task frequencies are converted into time shares to evaluate the proportion of working time that occupation i allocates to task k. Second, time shares are scaled by employment counts to calculate FTE headcounts at the task levels. Third, tasks are aggregated into cross-occupational DWAs to enable standardized comparisons across occupations. Uncertainty is propagated at each stage to provide confidence intervals of workforce estimates.

3.6.1. Occupation-level Time Shares

Let $\hat{\mu}_{i,k}$ denote the expected annual number of occurrences of task k in occupation i, let $d_{i,k}$ be a fixed task duration (hours per occurrence), and let K_i be the set of task indices within occupation i. Define the occupation-level total number of task occurrences:

$$M_i \;=\; \sum_{j \in K_i} \hat{\mu}_{i,j}\, d_{i,j}. \tag{11}$$

The time share $\pi_{i,k}$ that occupation i allocates to task k is

$$\pi_{i,k} \;=\; \frac{\hat{\mu}_{i,k}\, d_{i,k}}{M_i}, \qquad \text{where} \quad \sum_{k \in K_i} \pi_{i,k} = 1. \tag{12}$$

$\pi_{i,k}$ represents the expected percentage of time someone doing job i will spend on task k. In addition, we know $\mathrm{Var}(\hat{\mu}_{i,k})$ and we assume that $\hat{\mu}_{i,k}$ is asymptotically normal (van der Vaart 2000). Under this assumption, we propagate uncertainty using the delta method, following standard variance approximations for ratio estimators in survey sampling (Särndal, Swensson, and Wretman 1992). The resulting approximation is:

$$\frac{\partial \pi_{i,k}}{\partial \hat{\mu}_{i,k}} = \frac{d_{i,k}}{M_i}\big(1-\pi_{i,k}\big), \qquad \frac{\partial \pi_{i,k}}{\partial \hat{\mu}_{i,j}} = -\frac{\pi_{i,k}\, d_{i,j}}{M_i} \quad (j \neq k), \tag{13}$$

and we get:

$$\mathrm{Var}(\pi_{i,k}) \;\approx\; \frac{1}{\big(M_i\big)^2}\left[d_{i,k}^2\big(1-\pi_{i,k}\big)^2\,\mathrm{Var}(\hat{\mu}_{i,k}) \;+\; \pi_{i,k}^2 \sum_{\substack{j\in K_i \\ j\neq k}} d_{i,j}^2\,\mathrm{Var}(\hat{\mu}_{i,j}) \right]. \tag{14}$$

While the general expression holds, we follow the standard simplifying assumption that all tasks within an occupation have equal duration per occurrence and set $d_{i,j} = 1$ (Tomlinson et al. 2025; Martin and Monahan 2022). This assumption is required because O*NET reports task frequencies but not the time spent per execution. Importantly, this does not remove task heterogeneity: differences in total task time remain captured through the annualization vector w (6), which places greater weight on high-frequency categories (e.g., hourly) than on low-frequency ones (e.g., yearly) (Martin and Monahan 2022). Under this assumption, (11) reduces to $M_i = \sum_{j\in K_i} \hat{\mu}_{i,j}$ and, by independence, $\mathrm{Var}(M_i) = \sum_{j\in K_i} \mathrm{Var}(\hat{\mu}_{i,j})$. Equation (14) then simplifies to:

$$\mathrm{Var}(\pi_{i,k}) \;\approx\; \frac{1}{M_i^2}\Big[\big(1-2\pi_{i,k}\big)\,\mathrm{Var}(\hat{\mu}_{i,k}) \;+\; \pi_{i,k}^2\,\mathrm{Var}(M_i)\Big], \tag{15}$$

and we show how to evaluate the proportion of time that occupation i on task k, including a measure of uncertainty.

3.6.2. Workforce-level FTE Estimation

Having established occupation–task time shares $\pi_{i,k}$ and their associated uncertainty, we scale these proportions by employment levels to obtain task-level FTE counts. Let N_i denote the number of workers in occupation i and $\mathrm{SE}(N_i)$ its sampling standard error. The estimated FTE devoted to task k within occupation i is

$$\widehat{\mathrm{FTE}}_{i,k} = \pi_{i,k}\, N_i, \tag{16}$$

which satisfies $\sum_{k=1}^{K_i} \widehat{\mathrm{FTE}}_{i,k} = N_i$ by construction. To quantify uncertainty, we treat $\pi_{i,k}$ and N_i as independent random variables and apply the exact variance formula for the product of independent random variables (Goodman 1960). Writing $\mathrm{Var}(N_i) = \mathrm{SE}(N_i)^2$, this yields

$$\mathrm{Var}\Big(\widehat{\mathrm{FTE}}_{i,k}\Big) = N_i^2\,\mathrm{Var}(\pi_{i,k}) \;+\; \pi_{i,k}^2\,\mathrm{Var}(N_i) \;+\; \mathrm{Var}(\pi_{i,k})\,\mathrm{Var}(N_i). \tag{17}$$

When employment counts are treated as fixed, $\mathrm{SE}(N_i) \approx 0$ and the terms involving $\mathrm{Var}(N_i)$ vanish. The resulting estimate of task-level FTE is reported with a $(1-\alpha)$ confidence interval:

$$\widehat{\mathrm{FTE}}_{i,k} \;\pm\; z_{1-\alpha/2}\sqrt{\mathrm{Var}\Big(\widehat{\mathrm{FTE}}_{i,k}\Big)}, \tag{18}$$

truncated to the interval $[0, \sum_i N_i]$ to enforce natural bounds.

3.6.3. Aggregating Task FTEs

A key element of O*NET is the tasks to DWA assignments, which link occupation-specific tasks to shared DWAs through a many-to-many relationship. This structure enables an inter-occupation workforce-level analysis of FTEs since DWAs are shared between occupations. However, aggregating task-level FTEs to DWA FTEs introduces potential double-counting. To address this, we assume a uniform distribution of FTEs across linked DWAs and implement double-counting safeguards.

Let $\mathcal{M}_{i,k}$ denote the set of DWAs m linked to task k in occupation i. For each $m \in \mathcal{M}_{i,k}$, the FTE contribution from task k in occupation i to DWA m is allocated as:

$$\widehat{\mathrm{FTE}}_{i,k,m} = \frac{\widehat{\mathrm{FTE}}_{i,k}}{|\mathcal{M}_{i,k}|}, \tag{19}$$

where $|\mathcal{M}_{i,k}| \geq 1$ (in the special case when $|\mathcal{M}_{i,k}| = 1$, the full $\widehat{\mathrm{FTE}}_{i,k}$ is assigned to the single DWA). The total FTE for a given DWA m is obtained by summing over all contributing task-occupation pairs:

$$\widehat{\mathrm{FTE}}_m = \sum_i \sum_{\substack{k \\ m \in \mathcal{M}_{i,k}}} \widehat{\mathrm{FTE}}_{i,k,m}. \tag{20}$$

The mappings $|\mathcal{M}_{i,k}|$ are fixed from the O*NET dataset and thus introduce no additional uncertainty. Under the assumptions of independence across tasks within an occupation and across occupations (as in prior variance propagations), the variance of each contribution simplifies to:

$$\mathrm{Var}(\widehat{\mathrm{FTE}}_{i,k,m}) = \left(\frac{1}{|\mathcal{M}_{i,k}|}\right)^2 \mathrm{Var}(\widehat{\mathrm{FTE}}_{i,k}), \tag{21}$$

using the variance derived for $\widehat{\mathrm{FTE}}_{i,k}$ in previous sections. The total variance for $\widehat{\mathrm{FTE}}_m$ is then:

$$\mathrm{Var}(\widehat{\mathrm{FTE}}_m) = \sum_i \sum_{\substack{k \\ m \in \mathcal{M}_{i,k}}} \mathrm{Var}(\widehat{\mathrm{FTE}}_{i,k,m}), \tag{22}$$

with a corresponding $(1 - \alpha)$ confidence interval given by:

$$\widehat{\mathrm{FTE}}_m \pm z_{1-\alpha/2} \sqrt{\mathrm{Var}(\widehat{\mathrm{FTE}}_m)}, \tag{23}$$

truncated to $[0, \sum_i N_i]$ to respect natural bounds. This aggregation transforms task-level estimates into DWA-level workforce FTEs. Since many DWAs are shared among workers in the same industry, this enables direct comparison of activities across occupations in the intralogistics sector. Provided with uncertainty measures, this enables workforce planning and AI-HRC design by reporting the estimated FTE count performing each activity.

4. Results

This section provides a structured overview of our findings. The first subsection presents results on the data cleaning procedure and descriptive statistics of the intralogistics sector. The subsequent subsection evaluates the performance and diagnostics of the covariance completion methodology. The final subsection combines DWA-level FTE estimates with AI exposure scores to characterize employment concentration across DWAs and quantify the share of intralogistics workers in highly exposed activities.

4.1. *Results - Data Cleaning and Descriptive Statistics*

The data cleaning methodology described in Section 3.4 is applied to the intralogistics sector. Occupations are initially filtered to keep only the "Transportation and Warehousing" sector ones. Because this classification is broader than intralogistics, the analysis focuses on occupations within transportation and warehousing that involve internal material handling, coordination, and operational management within logistics facilities, excluding external transport and last-mile delivery. The selected occupations include both manual and managerial roles, such as *stockers*, *machine operators*, *logisticians*, and *supply-chain managers* (Appendix F).

Table 1.: Intralogistics dataset attrition by step: occupations and tasks.

	Occupations		**Tasks**	
Step	**Count**	**Cum. loss (%)**	**Count**	**Cum. loss (%)**
Initial intralogistics filter	20	—	408	—
After Task Ratings cleaning	14	30.00	219	46.32
After BLS 2024 match	11	45.00	167	59.07
Final vs. initial	11	45.00	167	59.07

The initial filtering of the intralogistics subset yields 20 Standard Occupational Classification (SOC) and 408 unique task statements. The Task Ratings cleaning procedures produce 14 occupations and 219 tasks. Alignment with the 2024 BLS naming convention further reduces the subset to 11 occupations and 167 tasks mapped to 98 DWAs. Table 1 summarizes the retention statistics.

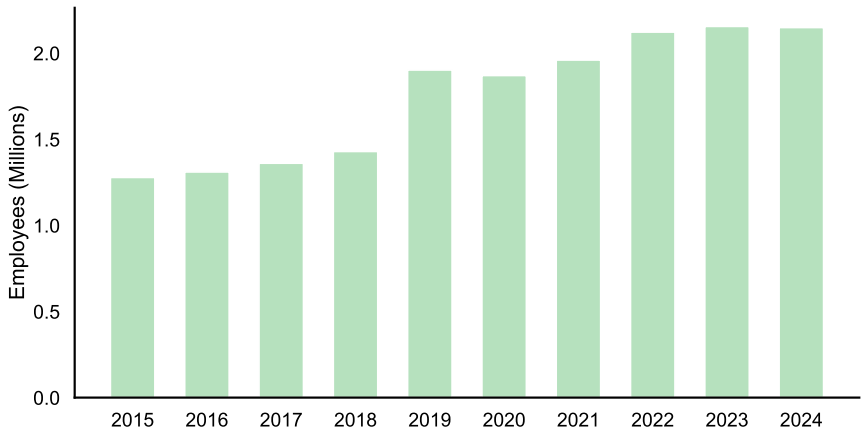

Figure 2.: Total employment in the filtered intralogistics sector, USA (2015–2024). Source: U.S. Bureau of Labor Statistics (2024).

Figure 2 illustrates the evolution of total employment in the sector from 2015 to 2024, based on data from the U.S. Bureau of Labor Statistics (2024). Employment increased steadily until 2021 and then plateaued. The spike in 2019 is attributable to an O*NET occupational reclassification (see Appendix G). Since 2022, the intralogistics sector has consistently employed approximately 2.2 million workers in the United States, representing about 1.4 percent of the total American workforce.

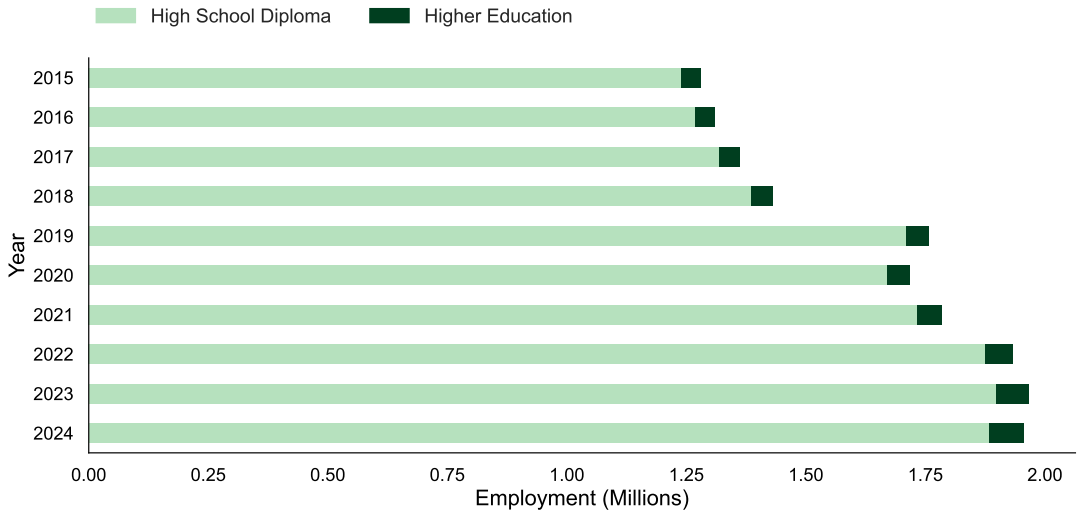

Figure 3.: Filtered intralogistics employment by education level (2015–2024). Source: U.S. Bureau of Labor Statistics (2024) & National Center for O*NET Development (2024).

One interesting insight is the breakdown of the U.S. intralogistics workforce by education level. Roles requiring a higher education degree grew rapidly, increasing from approximately 40,000 in 2015 to over 70,000 in 2024. In contrast, employment among workers with only a high school diploma grew from 1.34 million to about 2.01 million, reflecting slower relative growth rates of 5.8 percent and 4.8 percent compound annual growth rate, respectively. This percentage-point difference is primarily attributable to an accelerated growth in recent years. These patterns support a gradual but persistent

shift toward higher educational attainment within the intralogistics sector.

4.2. *Results - Covariance Completion Diagnostics and Robustness*

This section presents diagnostics for the covariance matrices, focusing on feasibility, geometric conformity, and sensitivity. Results are provided for both the complete cleaned dataset ($N = 11{,}834$) and the subset of intralogistics DWAs ($N = 167$).

(1) The assessment of covariance feasibility involves verifying the numerical and geometric validity of the completed covariance matrices. All task covariance matrices in the datasets are positive semi-definite. The maximum tangent residual is less than 10^{-10} indicating that the covariance matrices reside on the feasible simplex. Detailed results are provided in Appendix E.1.
(2) For each task, the estimated variance of the annual occurrence, derived from the completed covariance matrix, is compared to its feasible range. This range is determined through convex optimization with identical marginal constraints, resulting in task-specific lower and upper variance bounds. Each task's realized variance is normalized to a value between zero and one within these bounds. The mean normalized position for the intralogistics sector is 0.39 with a standard deviation of 0.12. No mean normalized position outlier was present for the intralogistics sector's tasks. In contrast, 177 outliers (1.5 percent) were identified in the full dataset of 11,834 tasks. Detailed results are available in Appendix E.2.
(3) The sensitivity of the covariance matrix to minor changes in estimates is evaluated by perturbing the target marginal variances by plus or minus 5 percent and recalculating the full covariance completion for each task. The symmetrized relative change in variance is then measured. All intralogistics tasks exhibit stable and well-conditioned behavior, as modest survey errors do not result in statistically significant changes in task variance. In comparison, the full dataset has 1 (0.01 percent) tasks failing directional checks. Detailed results are provided in Appendix E.3.

Diagnostics for the entire dataset identified 177 unique tasks (1.5 percent), where the covariance matrix did not satisfy at least one criterion. The small proportion supports the use of covariance completion to reconstruct matrices from marginal standard errors for task frequency estimates. No covariance matrix was removed for intralogistics tasks, yielding 167 appropriately behaved task matrices. These diagnostics support the reliability of our method used to calculate the FTE of the task for the intralogistics workforce.

4.3. *Results - Exposure to AI from Intralogistics Workers*

The main findings are now discussed in detail. Figure 4 illustrates estimated employment concentration across the ten largest DWAs and their associated uncertainty in the intralogistics sector. The columns indicate mean FTE estimates in thousands, and the error bars represent 90 percent confidence intervals. For example, the DWA *"move materials, equipment, or supplies"* is estimated to employ between 200,000 and 340,000 FTEs. Notably, employment is concentrated in a limited number of DWAs, which correspond to the most common operational roles. For instance, the ten largest DWAs account for 54 percent of the total intralogistics workforce (in terms of FTEs). This concentration reflects the sector's DWA structure, where recurrent operations

such as moving materials or supplies, monitoring cargo area conditions, and recording production data account for the largest FTE counts. The top ten DWAs by FTE count are listed in Appendix H.

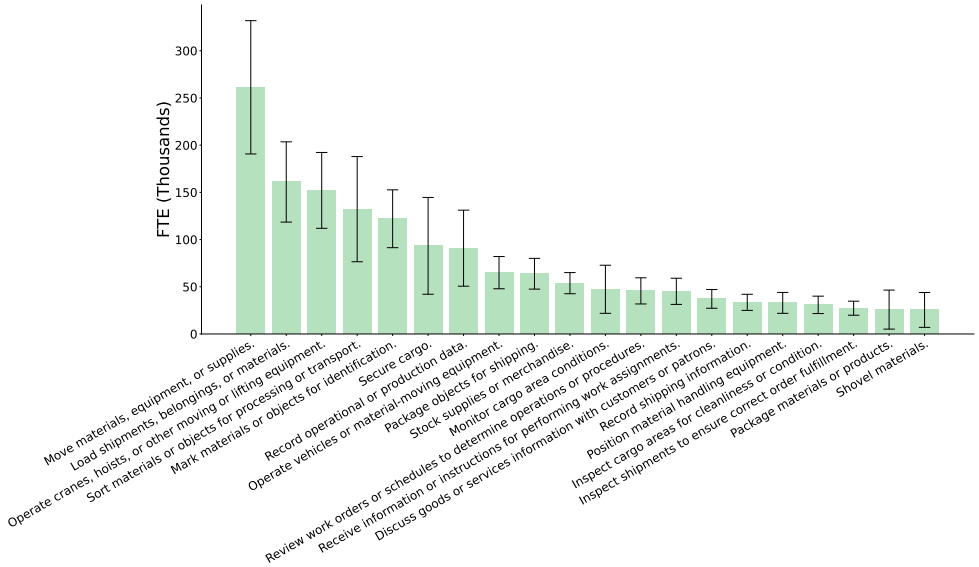

Figure 4.: Estimated FTEs by DWA Title with 90 percent Confidence Intervals.

Figure 5 associates each DWA to its AI exposure score. The x-axis represents exposure to AI-driven automation, and the y-axis indicates FTE for each DWA. Each point represents a DWA's exposure and its mean FTE estimate with a 90 percent confidence interval. The shaded area depicts the cumulative FTE distribution. The 85th percentile of the automation score identifies the most exposed DWAs which are highlighted as red crosses (Bouquet, Kaboli, and Sheffi 2025). This categorization reveals that the equivalent of 16.4 percent, or approximately 366,000 FTEs of the intralogistics workforce may be affected by the implementation of AI-HRC. The list of DWAs, their FTEs and confidence intervals are depicted in Table 2.

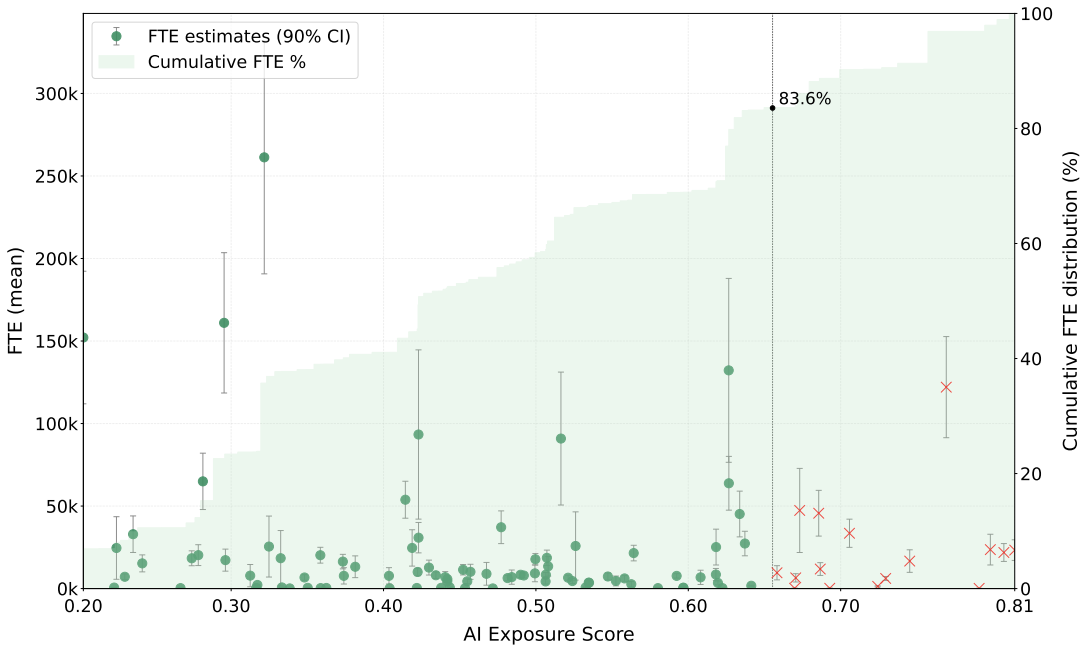

Figure 5.: FTE Mean vs. AI Exposure with 90 percent Confidence Intervals and Cumulative Distribution.

Within the high-exposure DWAs in Table 2, three tasks stand out by FTE magnitude: (i) *Mark materials or objects for identification*, (ii) *Monitor cargo area conditions*, and (iii) *Review work orders or schedules to determine operations or procedures*. Although these activities differ in function, they share factors that make them particularly susceptible to partial automation; namely, repetitive visual or procedural inspection and the integration of digital work-order systems.

Table 2.: DWAs above 85th percentile AI-HRC exposure, ordered by FTE.

DWA Title	**FTE Estimate**	**SE**
Mark materials or objects for identification	122,047	18,623
Monitor cargo area conditions	47,346	15,487
Review work orders or schedules to determine operations or procedures	45,661	8,424
Record shipping information	33,522	5,191
Attach identification information to products, items, or containers	23,610	5,659
Receive shipments	23,314	3,798
Read work orders to determine material or setup requirements	21,862	3,283
Prepare documentation for contracts, transactions, or regulatory compliance	16,669	4,149
Verify information or specifications	11,842	2,349
Analyze shipping information to make routing decisions	9,579	2,611
Store records or related materials	6,502	1,531
Measure product or material dimensions	6,140	723
Record production information	1,352	202
Prepare products for testing	198	71
Collect samples of materials or products for testing	198	71
Prepare informational or reference materials	128	51
Total	**366,958**	**29,087**

These results collectively indicate that while employment remains concentrated in a few dominant DWAs, the subset of tasks positioned at the higher end of the AI ex-

posure distribution tends to include both operational and supervisory functions. This heterogeneity suggests that the next stage of technological transformation in intralogistics may not involve the replacement of full occupations, but rather the augmentation of specific tasks. Consequently, precise strategic workforce planning is necessitated at the DWA level to effectively manage this transition.

5. Discussion

The uncertainty-aware approach developed in this study provides an understanding of the range of labor distribution across tasks. The following discussion connects this new methodology to three operational domains: strategic workforce planning, AI-HRC design, and resilience under uncertainty.

5.1. *Functional Areas and Task-Level Reallocation for AI-HRC*

Using our results, DWAs can be grouped into three domains ripe for AI-HRC: administrative and coordination tasks, material handling, and monitoring.

In administrative and coordination functions, enterprise resource planning systems already streamline documentation, data transfer, and scheduling. Generative AI assistants now extend these systems by automatically summarizing orders and drafting reports, enhancing efficiency in what were once manual clerical and supervisory workflows (Pennathur et al. 2024).

In material handling and storage, automation technologies such as AMRs, RMFS, and robotic process automation embody the physical side of AI integration. Computer vision, path planning, and digital twin models enable robots to navigate, transport goods, and coordinate warehouse operations dynamically (Fragapane et al. 2021; Li et al. 2023; Boysen and De Koster 2025). In these contexts, automation can replace human labor. Tasks once done manually by workers can be handled by machines, while oversight, exception management, and safety remain with human operators. These supervisory roles entail distinct responsibilities and may be filled by personnel other than those who performed manual labor.

Finally, in monitoring and inspection, perception-driven AI and sensor networks are automating quality control and environmental supervision. Vision-based defect detection, predictive maintenance, and anomaly detection systems continuously collect and interpret data, shifting the human role from routine checking toward diagnostic and decision-making functions (Liu et al. 2022; Li et al. 2023).

Overall, these three domains illustrate how AI-HRC transforms intralogistics into a hybrid ecosystem where automation enhances precision and productivity while humans are responsible for oversight and exception handling. In this context, targeted, task-specific interventions allow for more precise allocation of training resources and can improve workforce placement outcomes as AI-HRC displaces human labor. A detailed section with the selected DWAs and ready-to-deploy AI-HRC technologies is presented in Appendix I.

5.2. *Strategic Workforce Planning and Real-Time Decision Making*

Traditional workforce planning typically employs a top-down approach, with managers estimating headcount based on projected revenue and demand (Cantrell and Klosk 2025). Instead of viewing workforce planning as a periodic activity, the proposed framework supports continuous, real-time monitoring of task allocations between human workers and robots. The task-level model can be updated as employees are hired, reassigned, or their employment ends. When integrated with data on AI-driven robotic innovations, this approach transforms AI-HRC and workforce planning from a static process into a dynamic organizational intelligence system (Teece, Pisano, and Shuen 1992; Xu, Liu, and Montreuil 2025).

This dynamic approach facilitates scenario-based staffing strategies, allowing organizations to simulate the effects of AI-HRC and directly connect automation to workforce impact. These simulations provide insights into operational risk and enable the evaluation of workforce strategies for AI-HRC design. For instance, the largest intralogistics task, *"move materials, equipment, or supplies"*, has a 90 percent confidence interval of approximately 200,000 and 340,000 FTEs, supporting multiple staffing scenarios grounded in empirical data. A conservative AI-HRC scenario with a 90/10 human-to-robot ratio would affect between 20,000 and 34,000 FTEs, while an aggressive 40/60 split would reduce the workforce by at least 120,000 FTEs. Even limited automation can significantly affect workforce size and wage bill costs.

As human labor may remain less expensive than AI-HRC, this research can help identify the inflection point at which automation becomes economically viable for the organization and the expected amortization period from AI-HRC investment. Scenario analysis provides a proactive, data-driven basis for workforce planning and clarifies accountability for workforce and investment decisions.

5.3. *Employee Retention and Supply-Chain Resilience*

The intralogistics sector faces substantial workforce volatility, with an annual turnover rate of 49 percent, nearly three times the U.S. average and among the highest across all sectors (U.S. Bureau of Labor Statistics 2024). This labor instability is a major supply chain vulnerability, as disruptions to human capital can cause operational delays and performance losses (Nagurney and Ermagun 2022). AI-HRC helps mitigate these risks by improving working conditions in hazardous environments, reallocating personnel from repetitive tasks to higher-value activities, and increasing employee tenure and job satisfaction (Baltrusch et al. 2021).

Identifying workers in high-exposure DWAs who are suitable for upskilling and internal mobility prior to installing automation preserves firm-specific knowledge and stabilizes the workforce. The proposed framework supports AI-HRC planning by enabling organizations to quantify workforce impacts before deployment of automation and to invest in targeted capability development. From this perspective, AI-HRC represents an opportunity to enhance operations and invest in structured upskilling and career advancement. Such investments generate both financial and strategic benefits. Internal workforce development reduces recruiting costs and knowledge attrition, whereas external hiring is generally more expensive and associated with lower initial performance and higher turnover compared to internal movers (Bidwell 2011). Furthermore, organizational career growth is a strong predictor of employee retention, thereby reducing workforce volatility (Rani, Sood, and Chaurasia 2024).

In summary, proactive workforce planning strengthens supply chain resilience by

promoting workforce continuity and delivering financial advantages compared to external hiring.

6. Conclusion, Managerial Implications, and Future Research

We present a stochastic, task-based workforce framework that converts O*NET task frequencies into annualized occurrences, completes covariances under simplex constraints via a semidefinite program, and propagates uncertainty to task- and DWA-level full-time-equivalent (FTE) estimates. The method yields interpretable confidence intervals and detailed employment maps that show where intralogistics labor time is concentrated. Results indicate that the most AI-exposed DWAs span both operational and supervisory activities.

Three findings stand out. First, intralogistics employment is concentrated in a relatively small set of DWAs, so a focused subset of activities accounts for a large share of the sector's labor time. Second, the top 15 percent of DWAs suitable for AI-HRC encompasses several hundred thousand FTEs, indicating that even conservative adoption paths will affect workforce allocations at scale. Third, the most exposed DWAs span operational as well as coordination and administrative functions, so that near-term change may be multifaceted and will require careful, task-level workforce planning.

RQ1: Can we construct task-level full-time equivalent (FTE) estimates and characterize their statistical uncertainty?

Our findings show that a stochastic workforce framework can systematically estimate FTE requirements per task by modeling task frequencies as probability vectors on the simplex, completing their covariance structure, and propagating this uncertainty to task-level FTEs. The approach makes its assumptions explicit and evaluates mathematical consistency through feasibility, geometry, and sensitivity diagnostics. The resulting confidence intervals translate task frequency data from surveys into a range of workforce estimates, supporting more granular planning by identifying where labor time concentrates and how robust these estimates are to sampling and modeling uncertainty.

RQ2: How can task-level FTE estimates combined with AI exposure metrics inform strategic AI-HRC design and workforce planning?

Applied to the intralogistics sector, the stochastic workforce framework provides FTE estimates per task and per DWA, which can be combined with task-level automation exposure metrics to inform AI-HRC design. By linking FTE estimates to exposure indices, the framework identifies DWAs with both substantial labor demand and high suitability for AI-HRC, thereby quantifying the number of workers whose tasks are most exposed to reallocation or elimination under different automation scenarios. This supports scenario-based workforce planning (e.g., conservative versus aggressive AI-HRC adoption), allows managers to anticipate where headcount reductions or role redesign are most likely to be required, and highlights priority areas for upskilling and redeployment. In turn, such targeted planning can improve employee retention and, ultimately, enhance intralogistics and supply chain resilience.

6.1. *Managerial Implications*

- **AI-HRC design.** Combining FTE estimates with automation exposure metrics enables targeted prioritization of AI-HRC initiatives. High-FTE, high-exposure DWAs can be treated as priority candidates for intervention, whether through deploying collaborative robots, implementing advanced decision-support tools, or redesigning workflows. Conversely, DWAs with lower exposure or high uncertainty in FTE estimates may warrant more cautious experimentation. Aligning AI-HRC investments with this task-level analytics can help ensure that resources are focused on activities where they are most likely to generate productivity gains.
- **Strategic workforce planning.** Task-level FTE estimates and their confidence intervals support scenario-based staffing decisions. By simulating conservative and aggressive AI-HRC adoption paths (for example, shifting a fraction of task time to automated systems), managers can quantify how many FTEs are affected under each scenario and assess the risk associated with different automation strategies. This moves workforce planning beyond deterministic headcount projections toward uncertainty-aware, task-driven analyses.
- **Resilience under uncertainty.** The framework supports proactive talent management. By identifying the DWAs where AI-HRC is most likely to change work content, firms can design upskilling and reskilling pathways that anticipate task reallocation rather than reacting to it ex post. In this way, AI-HRC can be positioned as an instrument for workforce stabilization and long-term operational robustness.

6.2. *Future Research*

Several avenues for future research emerge from this study. First, relaxing the assumptions of equal task durations within occupations and uniform split of task time across linked DWAs using empirical data would allow richer heterogeneity as well as more precise task–DWA mappings. Second, the current analysis treats task frequencies and employment levels as cross-sectional. Extending the framework to panel data would enable analysis of dynamic adjustments in task composition and FTE allocations around AI-HRC adoption, including quasi-experimental designs. Third, combining the stochastic workforce framework with wage data would permit evaluation of how AI-HRC affects the distribution of labor income across DWAs, not just quantities of labor. Finally, the automation exposure metrics are derived from external text corpora; validating them with direct measures of technology deployment would strengthen the framework and better support the design of operationally effective and socially sustainable AI-HRC systems.

Disclosure of interest

The authors report no conflict of interest.

Data availability statement

The authors confirm that the data supporting the findings of this study are available. The experiments were conducted in a reproducible manner, and the source code, as well as data are available on the GitHub repository: `https://github.com/Pierre-Bouquet/Measuring-the-Uncertain-Task-Content-of-Work`.

References

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33 (2): 3–30.

Aitchison, John. 1982. "The Statistical Analysis of Compositional Data." *Journal of the Royal Statistical Society: Series B (Methodological)* 44 (2): 139–160.

Aleotti, Jacopo, Alberto Baldassarri, Marcello Bonfè, Marco Carricato, Davide Chiaravalli, Roberto Di Leva, Cesare Fantuzzi, et al. 2021. "Toward future automatic warehouses: An autonomous depalletizing system based on mobile manipulation and 3d perception." *Applied Sciences* 11 (13): 5959.

Autor, David, Caroline Chin, Anna Salomons, and Bryan Seegmiller. 2024. "New Frontiers: The Origins and Content of New Work, 1940–2018." *Quarterly Journal of Economics* 139 (3): 1399–1465.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association* jvaf023.

Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29 (3): 3–30.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market." *American Economic Review* 103 (5): 1553–1597.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4): 1279–1333.

Baltrusch, Saskia, Frank Krause, Aijse de Vries, and Michiel P. de Looze. 2021. "What about the Human in Human–Robot Collaboration? A literature review on HRC's effects on aspects of job quality." *Ergonomics* 65 (9): 1–35.

Bevington, Philip R., and D. Keith Robinson. 2003. *Data Reduction and Error Analysis for the Physical Sciences*. 3rd ed. McGraw-Hill.

Bidwell, Matthew. 2011. "Paying more to get less: The effects of external hiring versus internal mobility." *Administrative Science Quarterly* 56 (3): 369–407.

Bouquet, Pierre, Amin Kaboli, and Yossi Sheffi. 2025. "News Sentiment as a Dynamic Predictor of Job Automation Risk." Working paper.

Boyd, Stephen, and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Boysen, Nils, and René De Koster. 2025. "50 Years of Warehousing Research—An Operations Research Perspective." *European Journal of Operational Research* 320 (2): 449–464.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics* 140 (2): 889–942.

Brynjolfsson, Erik, and Tom Mitchell. 2017. "What Can Machine Learning Do? Workforce Implications." *Science* 358 (6370): 1530–1534.

Cai, Min, Gen Wang, Xinggang Luo, and Xueqi Xu. 2025. "Task Allocation of Human–Robot Collaborative Assembly Line Considering Assembly Complexity and Workload Balance." *International Journal of Production Research* 63 (13): 4749–4775.

Cantrell, Sue, and Russell Klosk. 2025. "Is it time to break workforce planning out of its silo?" *Deloitte Insights* Accessed: 2025-11-01.

De Koster, René, and Debjit Roy. 2024. "Research: Warehouse and Logistics Automation

Works Better with Human Partners." *Harvard Business Review* .

De Koster, René, Debjit Roy, Yun Fong Lim, and Subodha Kumar. 2025. "Internet of Things in Intralogistics: Applications and Emerging Research." *Production and Operations Management* Forthcoming.

Dell'Acqua, Fabrizio, Saran Rajendran, Edward McFowland, Lisa Krayer, François Mollick, Hila Lifshitz-Assaf, Karim Lakhani, and Katherine Kellogg. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Working paper.

Dierdorff, Erich C., Jennifer J. Norton, Christina M. Gregory, Darwin Rivkin, and Phillip Lewis. 2009. *O*NET Analyst Occupational Skills Ratings: Procedures*. Technical Report. Raleigh, NC: National Center for O*NET Development.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. "GPTs are GPTs: Labor market impact potential of LLMs." *Science* 384 (6702): 1306–1308.

Fang, Yingjian, Qingxiao Wu, Sicong Li, Jian Guan, and Yunge Cui. 2024. "Enhanced YOLOv5s-based algorithm for industrial part detection." *Sensors* 24 (4): 1183.

Felten, Edward W., Manav Raj, and Robert Seamans. 2018. "A Method to Link Advances in Artificial Intelligence to Occupational Abilities." *AEA Papers and Proceedings* 108: 54–57.

Fragapane, Giuseppe, René de Koster, Fabio Sgarbossa, and Jan Ola Strandhagen. 2021. "Planning and Control of Autonomous Mobile Robots for Intralogistics: Literature Review and Research Agenda." *European Journal of Operational Research* 294 (2): 405–426.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114: 254–280.

Goodman, Leo A. 1960. "On the exact variance of products." *Journal of the American statistical association* 55 (292): 708–713.

Grussler, Christian, Anders Rantzer, and Pontus Giselsson. 2018. "Low-Rank Optimization with Convex Constraints." *IEEE Transactions on Automatic Control* 63 (11): 4000–4007.

Han, Song, Xiaoping Liu, Xing Han, Gang Wang, and Shaobo Wu. 2020. "Visual sorting of express parcels based on multi-task deep learning." *Sensors* 20 (23): 6785.

Horn, Roger A., and Charles R. Johnson. 2012. *Matrix Analysis*. 2nd ed. Cambridge University Press.

Huber, Peter J., and Elvezio M. Ronchetti. 2009. *Robust Statistics*. 2nd ed. Wiley.

Krnjaic, Aleksandar, Raul D Steleac, Jonathan D Thomas, Georgios Papoudakis, Lukas Schäfer, AW Keung To, Kuan-Ho Lao, et al. 2024. "Scalable multi-agent reinforcement learning for warehouse logistics with robotic and human co-workers." In *RSJ International Conference on Intelligent Robots and Systems (IROS)*, 677–684.

Lam, HY, Valerie Tang, and L Wong. 2024. "Raising logistics performance to new levels through digital transformation." *International Journal of Engineering Business Management* 16: 18479790241231730.

Lee, Donggun, Yong-Shin Kang, and Sang Do Noh. 2025. "Digital twin-driven deep reinforcement learning for real-time optimisation in dynamic AGV systems." *International Journal of Production Research* 1–19.

Li, Mingxing, Daqiang Guo, Ming Li, Ting Qu, and George Q. Huang. 2023. "Operation Twins: Production–Intralogistics Synchronisation in Industry 4.0." *International Journal of Production Research* 61 (15): 5193–5211.

Liu, Huwei, Li Zhou, Junhui Zhao, Fan Wang, Jianglong Yang, Kaibo Liang, and Zhaochan Li. 2022. "Deep-learning-based accurate identification of warehouse goods for robot picking operations." *Sustainability* 14 (13): 7781.

Liu, Yu, Shenle Pan, and Eric Ballot. 2024. "Unveiling the Potential of Digital Twins in Logistics and Supply Chain Management: Services, Capabilities, and Research Opportunities." *Digital Engineering* 3: 100025.

Ma, Benedict Jun, and María Jesús Saénz. 2025. "AI Can Improve How Humans and Robots Work: Warehouse Automation Is Booming as Companies Ramp Up Their Use of Robotics." *MIT Sloan Management Review* .

Mahmoudinazlou, Sasan, Abhay Sobhanan, Hadi Charkhgard, Ali Eshragh, and George Dunn. 2025. "Deep reinforcement learning for dynamic order picking in warehouse operations." *Computers & Operations Research* 107112.

Martin, Josh, and Ellys Monahan. 2022. *Developing a Method for Measuring Time Spent on Green Tasks*. Technical Report. Office for National Statistics.

MOSEK, ApS. 2025. "The MOSEK Python Fusion API manual." .

Nagurney, Anna, and Alireza Ermagun. 2022. "Resilience of supply chain networks to labor disruptions." *Findings* .

National Center for O*NET Development. 2024. "O*NET 28.3 Database." `https://www.onetcenter.org/db_releases.html`, May. Version 28.3.

Nourmohammadi, A., T. Arbaoui, M. Fathi, and A. Dolgui. 2025. "Balancing Human–Robot Collaborative Assembly Lines: A Constraint Programming Approach." *Computers & Industrial Engineering* 111154.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." *Science* 381 (6654): 187–192.

O*NET Resource Center. 2023. "Data Collection Information." `https://www.onetcenter.org/dataCollection.html`.

Park, Kyu Tae, Yoo Ho Son, and Sang Do Noh. 2021. "The architectural framework of a cyber physical logistics system for digital-twin-based supply chain control." *International Journal of Production Research* 59 (19): 5721–5742.

Pasparakis, Alexandros, Jelle de Vries, and René de Koster. 2023. "Assessing the Impact of Human–Robot Collaborative Order Picking Systems on Warehouse Workers." *International Journal of Production Research* 61 (22): 7776–7790.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. "The impact of ai on developer productivity: Evidence from github copilot." *arXiv preprint arXiv:2302.06590* .

Pennathur, Priyadarshini R, Valerie Boksa, Arunkumar Pennathur, Andrew Kusiak, and Beth A Livingston. 2024. "The future of office and administrative support occupations in the era of artificial intelligence: A state of the art review and future research directions." *International Journal of Industrial Ergonomics* 104: 103665.

Perotti, Sara, Luca Cannava, Jörg M. Ries, and Eric H. Grosse. 2025. "Reviewing and Conceptualising the Role of 4.0 Technologies for Sustainable Warehousing." *International Journal of Production Research* 63 (6): 2305–2337.

Peterson, Norman G., Michael D. Mumford, Walter C. Borman, P. Richard Jeanneret, Edwin A. Fleishman, and Kerry Y. Levin. 1997. *O*NET Final Technical Report. Volume I, II, III*. Technical Report. Salt Lake City, UT: American Institutes for Research. `https://files.eric.ed.gov/fulltext/ED455452.pdf`.

Rani, Nidhi, Gaytri Sood, and Shiv S. Chaurasia. 2024. "Exploring Organizational Career Growth: A Systematic Literature Review and Future Research Directions." *The International Journal of Human Resource Management* 35 (17): 4081–4117.

Rinderle-Ma, Stefanie, Karolin Winter, and Janik-Vasily Benzin. 2023. "Predictive compliance monitoring in process-aware information systems: State of the art, functionalities, research directions." *Information Systems* 115: 102210.

Saltelli, Andrea, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. 2004. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley.

Sanchez-Cubillo, Javier, Javier Del Ser, and José Luis Martin. 2024. "Toward fully automated inspection of critical assets supported by autonomous mobile robots, vision sensors, and artificial intelligence." *Sensors* 24 (12): 3721.

Särndal, Carl-Erik, Bo Swensson, and Jan H. Wretman. 1992. *Model Assisted Survey Sampling*. Springer-Verlag.

Song, Chenyang, and Zhipeng Wu. 2024. "Artificial intelligence-assisted RFID tag-integrated multi-sensor for quality assessment and sensing." *Sensors* 24 (6): 1813.

Sun, Yangjun, Ning Zhao, Longwei Tang, and Lei Luo. 2024. "Breaking the Limit on the Number of Robots Through Conflict-Free Scheduling in Robotic Mobile Fulfillment Systems." *IEEE Transactions on Automation Science and Engineering* 21 (5): 1456–1470.

Sundaram, Sarvesh, and Abe Zeid. 2023. "Artificial intelligence-based smart quality inspection for manufacturing." *Micromachines* 14 (3): 570.

Teece, David J, Gary P Pisano, and Amy Shuen. 1992. *Dynamic capabilities and strategic management*. Center for Research in Management, University of California, Berkeley.

Tomlinson, Kiran, Sonia Jaffe, Will Wang, Scott Counts, and Siddharth Suri. 2025. "Working with AI: measuring the applicability of generative AI to occupations." *arXiv preprint arXiv:2507.07935* .

Tsacoumis, Suzanne, and Chad H. Van Iddekinge. 2006. *A Comparison of Incumbent and Analyst Ratings of O*NET Skills*. Technical Report. O*NET Resource Center. `https://www.onetcenter.org/dl_files/SkillsComp.pdf`.

Tsang, YP, Valerie Tang, CH Wu, and Fuqing Li. 2024. "Unlocking the potential of robotic process automation for digital transformation in logistics and supply chain management." *Journal of Global Information Management (JGIM)* 32 (1): 1–24.

U.S. Bureau of Labor Statistics. 2024. "Occupational Employment and Wages — May 2024." `https://www.bls.gov/news.release/ocwage.htm`, Apr. News release summarizing May 2024 OEWS.

van der Vaart, Aad W. 2000. *Asymptotic Statistics*. Cambridge University Press.

Wan, Jianqiang, Sibo Song, Wenwen Yu, Yuliang Liu, Wenqing Cheng, Fei Huang, Xiang Bai, Cong Yao, and Zhibo Yang. 2024. "Omniparser: A unified framework for text spotting key information extraction and table recognition." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15641–15653.

Wolkowicz, Henry, Romesh Saigal, and Lieven Vandenberghe. 2000. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Springer.

Xu, Y., Y. Liu, and B. Montreuil. 2025. "Dynamic Workforce Scheduling and Relocation in Hyperconnected Parcel Logistic Hubs." In *Proceedings of the International Conference on Logistics and Supply Chain Management*, edited by K. Ellis, W. Ferrell, and J. Knapp, 245–268. Georgia Institute of Technology.

Yerlikaya, Mehmet Akif, and Feyzan Arıkan. 2024. "A novel framework for production planning and class-based storage location assignment: Multi-criteria classification approach." *Heliyon* 10 (18).

Yuan, Ming, and Yin Zhang. 2013. "Robust Covariance Matrix Completion." *Biometrika* 100 (4): 905–920.

Yue, Man-Chung, Yves Rychener, Daniel Kuhn, and Viet Anh Nguyen. 2024. "A geometric unification of distributionally robust covariance estimators: Shrinking the spectrum by inflating the ambiguity set." *arXiv preprint arXiv:2405.20124* .

Appendix A. Modeling Assumptions and Choices

The empirical strategy relies on a set of structural assumptions and pragmatic modeling choices that ensure tractability, internal consistency, and reproducibility. These assumptions underpin the estimation of task-time distributions, covariance structures, and FTE allocations.

- **Survey-stage independence.** The task relevance shares $\widehat{RT}_{i,k}$ and conditional frequency shares $\widehat{FT}_{i,k,r}$ are treated as statistically independent. This enables exact variance evaluation for unconditional category probabilities $\widehat{p}_{i,k,r}$ through the product-variance formula.
- **Aggregation independence.** Tasks are assumed independent within and across occupations, and the O*NET-derived task-time shares are assumed independent of BLS OEWS employment levels N_i when propagating uncertainty to FTE estimates.
- **Normal inference.** Interval estimates rely on normal approximations obtained from delta-method variance propagation, with truncation applied as needed to respect simplex or nonnegativity constraints.
- **Covariance completion.** The latent covariance matrix of the eight-category probability vector is recovered by selecting a positive semidefinite matrix that matches the observed O*NET marginal variances and satisfies the simplex tangent constraint ($\Sigma\mathbf{1} = 0$). Among feasible matrices, the Frobenius-norm–minimal solution is chosen as a parsimonious regularizer.
- **Annualization mapping.** Ordinal task-frequency categories are mapped to expected annual occurrence counts using a fixed midpoint vector w, providing a consistent and conservative annualization of task frequencies.
- **Equal per-occurrence duration.** Within an occupation, occurrences of a given task are assumed to have a common expected duration. Under this assumption, annualized occurrence counts proxy relative time shares across tasks.
- **Task–DWA allocation.** The O*NET crosswalk linking tasks to DWAs is treated as fixed. When a task maps to multiple DWAs, its FTE contribution is split evenly across linked DWAs to avoid double counting.
- **Numerical stabilization.** Small ε-regularization and renormalization are applied to near-zero probabilities or standard errors to maintain numerical feasibility. Solutions are accepted when positive semidefiniteness holds within standard solver tolerances.

Appendix B. Limitations

B.1. *Temporal Validity*

This analysis pairs an O*NET release with BLS employment data from the same period. As artificial intelligence capabilities, work organization, and task taxonomies evolve, exposure and FTE figures are most reliable for short- to medium-term planning. For operational purposes, regular re-estimation is advised as new O*NET releases and firm-level data become available. In rapidly changing environments, quarterly updates provide a sensible update frequency.

B.2. *Geographic and Facility Heterogeneity*

Sector totals, such as the approximately 2.2 million workers reported, represent economy-wide averages. These aggregate figures obscure variation by region, facility type, and technology stack. For facility-level workforce planning, the estimation pipeline should be rerun using local microdata. This reveals the site-specific distribution of DWAs and automation technology adoption. While macro-level patterns are useful for benchmarking, site-level decisions require localized data on tasks and operational constraints.

B.3. *AI Exposure as Proxy, not Prediction*

News-sentiment exposure scores reflect media coverage of AI-driven automation capabilities but do not guarantee business ROI, because of micro regulatory barriers, or labor-cost dynamics. High exposure indicates that the underlying capability is present and salient, not that deployment is imminent or cost-effective. Ground-truth validation against observed AMR/RMFS deployments, vendor announcements, and facility surveys constitutes an important avenue for future work to calibrate predictive value and refine decision thresholds.

B.4. *Task Dependence and Sequencing*

Within occupations, tasks often occur in sequences, such as receiving, marking, and moving. Shared time budgets can introduce dependence that extends beyond the scope of the current covariance completion. Diagnostics indicate a low effective rank for the completed covariance matrix, with about three to four of eight categories accounting for most variation. This suggests a limited number of shared tasks explain the majority of variation.

Collectively, these assumptions establish a transparent and manageable estimation pipeline. The methodology is reproducible and supports clear diagnostic evaluation. For critical assumptions, feasibility checks and sensitivity analyses are provided. Straightforward robustness extensions are outlined, including alternative covariance completions, midpoint stress tests, crosswalk weighting, and dependence-aware aggregation. These options can be applied by practitioners when site-specific data are available.

Appendix C. O*NET Data Examples

C.1. *O*NET structure and task hierarchy*

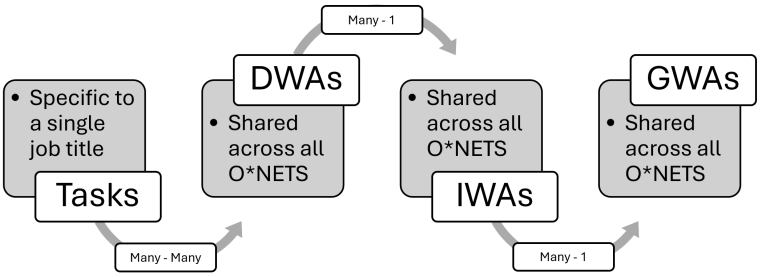

Figure C1.: O*NET task hierarchy.

C.2. *Survey mode and reported results*

Task data come from structured surveys administered to incumbent workers and occupation analysts (Dierdorff et al. 2009). The survey follows a two-stage response protocol. First, respondents indicate whether each task is relevant to their occupation using a binary relevance filter. Tasks marked as not relevant receive no further evaluation, and the resulting relevance measure serves as a proxy for whether a task is ever performed in that occupation (Martin and Monahan 2022). For tasks deemed relevant, respondents report the frequency of performance using a seven-point ordinal scale (O*NET Resource Center 2023)[1]. The published O*NET data report the share of respondents selecting each category. Illustrative examples appear in Appendix C.

For each occupation, O*NET aggregates incumbent responses and provides task-level statistics, including category proportions, sample sizes, and design-adjusted standard errors (Peterson et al. 1997). The reported 95 percent confidence intervals rely on normal-theory inference, justified by O*NET's stratified sampling design and the associated design-based Central Limit Theorem (Särndal, Swensson, and Wretman 1992). Consistent with prior work evaluating the sampling behavior of O*NET estimates using intraclass correlations (Tsacoumis and Van Iddekinge 2006), we assume approximate normality in the sampling distributions of the mean task-frequency estimators. We also assume independence between O*NET task shares and BLS employment counts, as the two systems draw on distinct sampling frames—worker-level surveys for O*NET and establishment-level surveys for BLS.

C.3. *Data table examples*

[1]The seven Frequency-of-Task (FT) categories are: 1 = Yearly or less, 2 = More than yearly, 3 = More than monthly, 4 = More than weekly, 5 = Daily, 6 = Several times daily, 7 = Hourly or more.

Table C1.: Example rows from the Task Ratings dataset for 11-1011.00 Chief Executives.

O*NET-SOC	Title	Task ID	Task	Scale ID	Scale Name	Cat.	Data Value	N	SE	LCI	UCI	Supp.	Date	Source
11-1011.00	Chief Executives	8823	Direct or...	FT	Frequency (1–7)	1	5.92	76	4.27	1.35	22.44	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	2	15.98	76	5.60	7.65	30.40	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	3	29.68	76	9.53	14.52	51.18	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	4	21.18	76	8.39	8.99	42.23	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	5	19.71	76	7.37	8.85	38.29	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	6	4.91	76	2.67	1.63	13.92	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	FT	Frequency (1–7)	7	2.63	76	2.22	0.48	13.21	N	08/2023	Incumbent
11-1011.00	Chief Executives	8823	Same as above	IM	Importance	–	4.52	75	0.11	4.31	4.74	N	08/2023	Incumbent

Table C2.: Example rows from the Tasks to DWA crosswalk for 11-1011.00 Chief Executives.

O*NET-SOC	Title	Task ID	Task	DWA ID	DWA Title	Date	Source
11-1011.00	Chief Executives	20461	Review and analyze legislation, laws, or public policy and recommend changes to promote or support interests of the general population or special groups.	4.A.2.a.4.I09.D03	Analyze impact of legal or regulatory changes.	07/2014	Analyst
11-1011.00	Chief Executives	20461	Same as above	4.A.4.b.6.I08.D04	Advise others on legal or regulatory compliance matters.	07/2014	Analyst
11-1011.00	Chief Executives	8823	Direct or coordinate an organization's financial or budget activities to fund operations, maximize investments, or increase efficiency.	4.A.4.b.4.I09.D02	Direct financial operations.	03/2014	Analyst
11-1011.00	Chief Executives	8824	Confer with board members, organization officials, or staff members to discuss issues, coordinate activities, or resolve problems.	4.A.4.a.2.I03.D14	Confer with organizational members to accomplish work activities.	03/2014	Analyst
11-1011.00	Chief Executives	8825	Analyze operations to evaluate performance of a company or its staff in meeting objectives or to determine areas of potential cost reduction, program improvement, or policy change.	4.A.2.a.4.I07.D09	Analyze data to assess operational or project effectiveness.	03/2014	Analyst
11-1011.00	Chief Executives	8826	Direct, plan, or implement policies, objectives, or activities of organizations or businesses to ensure continuing operations, to maximize returns on investments, or to increase productivity.	4.A.2.b.1.I09.D01	Implement organizational process or policy changes.	03/2014	Analyst

Table C3.: Example rows from the Transportation & Warehousing attachment. Percent shows share of occupation's workers employed in T&W.

Employed in T&W	Code	Occupation	Projected Growth 2024–2034	Projected Openings
99%	53-2031.00	Flight Attendants	Much faster than average	19,800
97%	53-4011.00	Locomotive Engineers	Slower than average	2,200
97%	53-1041.00	Aircraft Cargo Handling Supervisors	Faster than average	1,100
94%	53-2011.00	Airline Pilots, Copilots, and Flight Engineers	Average	11,700
92%	53-4022.00	Railroad Brake, Signal, and Switch Operators and Locomotive Firers	Slower than average	1,000
90%	43-5011.00	Cargo and Freight Agents	Much faster than average	8,800
90%	43-5011.01	Freight Forwarders	Much faster than average	8,800

Appendix D. Solver Settings and MOSEK Parameters

All semidefinite programs are solved with MOSEK's conic interior-point method under a strict tolerance profile. The main configuration is:

`solver_order` = (MOSEK), `optimiser_norm` = Frobenius, `accept_inaccurate` = **False**

Table D1.: MOSEK/IPM configuration used in all main experiments.

Parameter	Value	Purpose
`MSK_DPAR_INTPNT_CO_TOL_REL_GAP`	10^{-10}	Relative duality gap tolerance
`MSK_DPAR_INTPNT_CO_TOL_PFEAS`	10^{-10}	Primal feasibility tolerance
`MSK_DPAR_INTPNT_CO_TOL_DFEAS`	10^{-10}	Dual feasibility tolerance
`MSK_DPAR_INTPNT_CO_TOL_MU_RED`	10^{-10}	Barrier reduction (mu) tolerance
`MSK_IPAR_NUM_THREADS`	5	Parallelism
`MSK_IPAR_PRESOLVE_USE`	1	Enable presolve
`MSK_IPAR_INTPNT_SCALING`	1	Enable scaling

Appendix E. Covariance Diagnostics

E.1. *Feasibility and Simplex Geometry*

The numerical feasibility and compositional integrity of each task-level completed covariance matrix Σ were verified. For each task, feasibility diagnostics confirm that (i) the matrix is positive semidefinite (PSD), (ii) the simplex geometry implied by task-time shares is maintained, and (iii) the completed variances align with observed survey marginals.

E.1.1. Diagnostics Definitions

Let $\mathbf{1}$ be an all-ones vector of length $n = 8$ (frequency categories), and $v = (\mathrm{se}_p)^2$ the vector of design-based marginal variances. For each completed covariance Σ, the following metrics are computed:

$$\text{Tangent residual:} \quad r_\infty = \max\big(\|\Sigma\mathbf{1}\|_\infty, \|\mathbf{1}^\top\Sigma\|_\infty\big), \tag{E1}$$

$$\text{PSD check:} \quad \lambda_{\min}(\Sigma) \geq -10^{-8}, \tag{E2}$$

$$\text{Diagonal deviation:} \quad d_\infty = \|\mathrm{diag}(\Sigma) - v\|_\infty, \tag{E3}$$

$$\text{Uncertainty magnitude:} \quad \mathrm{trace}(\Sigma), \tag{E4}$$

$$\text{Effective rank:} \quad r_{\text{eff}} = \exp\Big(-\sum_j p_j \log p_j\Big), \qquad p_j = \lambda_j / \sum_\ell \lambda_\ell, \tag{E5}$$

In this context, λ_j denotes the eigenvalues of Σ, which are clipped at zero to enforce positive semidefiniteness. The entropy-based effective rank $r_{\text{eff}} \in [0, 7]$ quantifies the dimensionality of task-level uncertainty.

E.1.2. Results and Discussion

Table E1.: Covariance Completion Feasibility Diagnostics

	All Tasks (N = 11,834)		Intralogistics Tasks (N = 167)	
Metric	Value	Tasks Removed	Value	Tasks Removed
PSD Percentage	100 %	0	100 %	0
Max tangent residual	3.92×10^{-11}	0	3.28×10^{-11}	0
Max diagonal deviation	2.34×10^{-10}	0	1.94×10^{-10}	0
Trace (mean [SD])	0.021 [0.009]	n/a	0.022 [0.010]	n/a
Effective rank (mean [SD])	3.74 [1.02]	n/a	3.41 [0.89]	n/a

All 11,834 task-level matrices are positive semidefinite within the tolerance $\lambda_{\min} \geq -10^{-8}$, confirming that all quadratic forms used in downstream uncertainty propagation, including variance of annual task occurrences and FTE estimates, are well-defined. The maximum tangent residual $r_\infty \approx 10^{-11}$ for both the full and intralogistics subsets confirms that every covariance matrix lies within the simplex tangent space

($\Sigma\mathbf{1} = \mathbf{0}$), thereby preventing compositional leakage of variance outside the task-share simplex.

The diagonal deviations are negligible ($d_\infty < 3 \times 10^{-10}$), indicating near-perfect agreement with observed marginal sampling variances. The average trace $\text{trace}(\Sigma) = 0.021$; $SD = 0.009$ (all tasks) and $\text{trace}(\Sigma) = 0.022$; $SD = 0.01$ (intralogistics) quantifies the magnitude of task-level uncertainty inherited from O*NET marginals. Under the Frobenius-norm completion, a portion of variance is transferred to off-diagonal entries to satisfy PSD and the simplex constraint, so frequency categories within a task are not independent (as expected). The entropy-based effective rank values of 3.76 (all tasks) and 3.41 (intralogistics) out of eight categories confirm a low-dimensional structure, with most uncertainty concentrated in three to four dominant modes.

These diagnostics demonstrate that the covariance completions are numerically stable, geometrically valid, and compositionally consistent. The low effective rank facilitates efficient semidefinite optimization and supports interpretable uncertainty propagation. None of the diagnostic metrics are statistically significantly different between the full and intralogistics subsets, supporting the generalizability of the methodology beyond this sector.

E.2. *Functional Variance Bounds and Position*

The variance of an annual-occurrence functional is benchmarked against its feasible range for each task.

E.2.1. Diagnostics Definitions

Let $w \in \mathbb{R}^8$ denote the fixed annual-occurrence weights over the eight O*NET frequency categories, let $v = \text{Var}(p_{i,k})$ be the vector of design-based marginal variances, and let $\widehat{\Sigma}$ denote the completed task-level covariance matrix. The realized functional variance associated with task (i, k) is given by $w^\top \widehat{\Sigma}\, w$.

To assess whether this value is geometrically plausible given the marginal variances and the simplex constraint, we compute the feasible lower and upper bounds through the following convex programs, which impose the same constraints used during covariance completion—positive semidefiniteness, fixed diagonal entries, and tangency to the simplex:

$$\underline{V} = \min_{\Sigma \succeq 0} \; w^\top \Sigma\, w \quad \text{s.t. } \text{diag}(\Sigma) = v,\; \Sigma\mathbf{1} = 0, \tag{E6}$$

$$\overline{V} = \max_{\Sigma \succeq 0} \; w^\top \Sigma\, w \quad \text{s.t. } \text{diag}(\Sigma) = v,\; \Sigma\mathbf{1} = 0. \tag{E7}$$

The relative position of the completed functional variance within its task-specific feasible interval is then summarized by the normalized statistic

$$\text{pos} \;=\; \frac{w^\top \widehat{\Sigma}\, w \;-\; \underline{V}}{\overline{V} \;-\; \underline{V}} \;\in\; [0, 1]. \tag{E8}$$

Tasks whose normalized position lies near the boundaries of the feasible

set—specifically, pos < 0.05 or pos > 0.95—are flagged as extreme and removed from downstream analysis. This robustness filter excludes covariance structures that, while feasible, imply implausibly tight or excessively diffuse uncertainty relative to the geometric constraints of the probability simplex.

E.2.2. Results and Discussion

Table E2.: Variance Position and Outlier Summary

	All Tasks ($N = 11{,}834$)		Intralogistics Tasks ($N = 167$)	
Metric	Value	Tasks Removed	Value	Tasks Removed
Mean normalized position	0.41 [0.13]	n/a	0.39 [0.12]	n/a
Outlier (position < 0.05 / > 0.95)	1.5%	177	0%	0

Table E2 indicates that most tasks fall within the interior of their feasible variance ranges rather than at the extremes. The mean normalized position is 0.45 [SD 0.13] for all tasks and 0.39 [SD 0.12] for intralogistics tasks, suggesting no statistically significant difference between groups. Extreme positions are rare: 1.5 percent of all tasks (177 tasks) and none of intralogistics tasks are flagged and removed. This filtering prevents a few high-leverage cases from dominating uncertainty budgets and yields a parsimonious, decision-useful variance structure for downstream workforce planning and AI-HRC design. All bounds are computed with the same marginal and simplex constraints as the covariance completion, using semidefinite optimization. Solver settings are reported in Appendix D.

E.3. *Sensitivity Analysis*

We assess the stability of task-level functional variances to modest perturbations of the marginal targets.

E.3.1. Diagnostics Definitions

Let $v = (\mathrm{se}_p)^2$ denote the design-based marginal variances for the eight frequency categories and $w \in \mathbb{R}^8$ the annual-occurrence weights. For each task, with completed covariance $\widehat{\Sigma}$, the baseline functional variance is

$$V_0 \;=\; w^\top \widehat{\Sigma}\, w. \tag{E9}$$

We form perturbed diagonal targets $v^{(+)} = (1+\tau)\, v$ and $v^{(-)} = (1-\tau)\, v$ with $\tau = 0.05$, and recompute two completions by solving the same SDP used at baseline (PSD, fixed diagonal, simplex tangency $\Sigma \mathbf{1} = 0$). Let $V_+ = w^\top \widehat{\Sigma}^{(+)} w$ and $V_- = w^\top \widehat{\Sigma}^{(-)} w$. We then compute

$$\Delta_+ = V_+ - V_0, \qquad \Delta_- = V_- - V_0, \qquad \delta_\pm^{\mathrm{rel}} = \frac{\Delta_\pm}{\max(10^{-9},\, |V_0|)}. \tag{E10}$$

To diagnose linearity and symmetry, we use the symmetrized relative change

$$\text{symm\_rel} \;=\; \frac{\left|\Delta_+ + \Delta_-\right|}{\max(10^{-9},\, \max\{|\Delta_+|, |\Delta_-|\})}, \tag{E11}$$

and the spread:

$$\text{spread} = \frac{\max\{|\Delta_+|, |\Delta_-|\}}{\max(10^{-9}, \min\{|\Delta_+|, |\Delta_-|\})} \tag{E12}$$

Tasks are flagged if (i) symm_rel > 0.1 (nonlinear/asymmetric response), or if (ii) the directional relative changes leave the expected linear band $\delta_{\pm}^{\text{rel}} \notin [0.9\tau,\, 1.1\tau] = [0.045,\, 0.055]$.

E.3.2. Results and Discussion

Table E3.: Sensitivity Analysis (for ±5% perturbation)

	All Tasks (N = 11,834)		Intralogistics Tasks (N = 167)	
Metric	Value	Tasks Removed	Value	Tasks Removed
Symm. relative change	$5.49 \cdot 10^{-4}$ [$2.92 \cdot 10^{-3}$]	n/a	$1.67 \cdot 10^{-4}$[$3.26 \cdot 10^{-4}$]	n/a
High sensitivity (> 10%)	0.01%	1	0%	0
Delta rel. minus out of range	0.01%	1	0%	0
Delta rel. plus out of range	0.01%	1	0%	0
Total unique Tasks	0.01%	1	0%	0

Table E3 shows that responses are highly stable and the changes are not statistically significant different from 0. The mean symm_rel is small in both groups, 5.49×10^{-4} (SD 2.92×10^{-3}) overall and 1.67×10^{-4} (SD 3.26×10^{-4}) in intralogistics, indicating that V_+ and V_- move approximately anti-symmetrically around V_0 as expected under a first-order response. Only 0.01 percent of all tasks (1) exceeds the nonlinearity threshold, and none do so in intralogistics. Directional checks are likewise benign: 0.01 percent (1) of tasks fall outside the expected band for the negative perturbation and 0.01 percent (1) for the positive perturbation; again, none in intralogistics. Taking the union of all flags yields 0.01 percent (1) unique tasks overall.

These screens indicate that modest survey-error–like perturbations of the marginals do not induce large or erratic changes in functional variances. Intralogistics tasks are particularly well conditioned (no flags), reinforcing that the covariance completion behaves predictably under small input noise and that downstream planning metrics remain robust. Solver choices and tolerances match the baseline completion (see Appendix D).

Appendix F. Intralogistics Occupations Selected

Table F1.: Occupations selected for the intralogistics sector analysis.

Code	Occupation
53-7051.00	Industrial Truck and Tractor Operators
53-7011.00	Conveyor Operators and Tenders
53-7062.00	Laborers and Freight, Stock, and Material Movers, Hand
53-7062.04	Recycling and Reclamation Workers
53-1043.00	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo Handling Supervisors
53-1042.00	First-Line Supervisors of Helpers, Laborers, and Material Movers, Hand
53-1042.01	Recycling Coordinators
11-3071.00	Transportation, Storage, and Distribution Managers
11-3071.04	Supply Chain Managers
43-5111.00	Weighers, Measurers, Checkers, and Samplers, Recordkeeping
53-7041.00	Hoist and Winch Operators
53-7063.00	Machine Feeders and Offbearers
53-7064.00	Packers and Packagers, Hand
43-5071.00	Shipping, Receiving, and Inventory Clerks
53-7065.00	Stockers and Order Fillers
53-7021.00	Crane and Tower Operators
13-1081.00	Logisticians
13-1081.01	Logistics Engineers
13-1081.02	Logistics Analysts
53-6011.00	Bridge and Lock Tenders

Appendix G. Occupation Appearance

Table G1.: Job appearance and disappearance events in the intralogistics sector (2015–2024).

Year	Appeared Count	Appeared	Disappeared Count	Disappeared
2016	0		0	
2017	0		0	
2018	0		0	
2019	1	53-7065 – Stockers and Order Fillers	0	
2020	0		0	
2021	0		0	
2022	0		0	
2023	0		0	
2024	0		0	

Table G1 reports job appearance and disappearance events within the intralogistics sector from 2015 to 2024. The results show that a single occupational category, *Stockers and Order Fillers (SOC 53-7065)*, was introduced in 2019. No other entries or exits were observed during the period. Consequently, the apparent increase in employment observed in 2019 may be partially artificial, reflecting a classification update rather than a genuine expansion of the workforce. This adjustment should therefore be interpreted cautiously when analyzing longitudinal trends in employment or wage bill growth.

Appendix H. Top 10 Largest FTE DWAs

Table H1.: Top 10 DWAs by FTE Estimate (SD)

Rank	DWA Title	FTE Estimate (SD)
1	Move materials, equipment, or supplies.	261,300 (42,900)
2	Load shipments, belongings, or materials.	161,000 (25,850)
3	Operate cranes, hoists, or other moving or lifting equipment.	152,100 (24,400)
4	Sort materials or objects for processing or transport.	132,200 (33,850)
5	Mark materials or objects for identification.	122,050 (18,600)
6	Secure cargo.	93,350 (31,150)
7	Record operational or production data.	90,900 (24,450)
8	Operate vehicles or material-moving equipment.	64,950 (10,350)
9	Package objects for shipping.	63,800 (9,900)
10	Stock supplies or merchandise.	53,850 (6,800)

Note: Standard errors for our estimates are provided in parentheses. Values are rounded to the nearest 50.

Appendix I. Examples of AI-HRC in industry

We performed a literature review to examine the AI-HRC potential of each DWA. Below, we summarize our findings with DWAs organized in alphabetical order.

Analyze shipping information to make routing decisions.

Current research indicates that multi-agent reinforcement learning and automation integration significantly impact the task of analyzing shipping information for intralogistics routing. AI-driven routing automation reduces manual decision-making time by 40 percent while improving accuracy through real-time data analysis (Krnjaic et al. 2024). Furthermore, AMRs and automated guided vehicles (AGVs) embed routing algorithms to achieve 15–20 percent improvements in operational efficiency and reduce empty travel distances by 50 percent in warehouse distribution systems (Sun et al. 2024).

Attach identification information to products, items or containers.

Automated robotic marking systems integrated with AI-driven vision technologies substantially improve identification capabilities. Implementations utilizing deep learning object detection (such as YOLOv5s) and RFID systems have been shown to reduce identification errors by 25–40 percent and increase marking throughput by three to five times compared to manual methods, achieving an accuracy of 82.11 percent (Fang et al. 2024).

Collect samples of materials or products for testing.

Robotic automation significantly impacts sample collection tasks, offering substantial efficiency gains. Automated warehouse systems integrate robotic mobile fulfillment with real-time sample identification, achieving 93 percent accuracy in quality control (Liu et al. 2022).

Measure product or material dimensions.

Current implementation and research in intralogistics and AI-HRC significantly impact the task of measuring product or material dimensions. Key research highlights automated dimensional verification systems that reduce manual inspection time by 60–70 percent, while achieving measurement accuracy of 88.3–99.9 percent depending on implementation. As a result of such automation integration, dimensional accuracy improves from a manual baseline of 80 percent to an automated 96–99 percent (Sun et al. 2024).

Monitor cargo area conditions.

Current research demonstrates that AI and multi-robot systems improve anomaly detection, condition assessment, and asset protection through the use of vision sensors (Sanchez-Cubillo, Del Ser, and Martin 2024). Additionally, applied RPA solutions streamline temperature and status checks, yielding measurable gains in throughput,

fault reduction, and asset safety via automation integration (Lam, Tang, and Wong 2024).

Prepare documentation for contracts, transactions, or regulatory compliance.

The integration of RPA with machine learning has enabled the automation of critical administrative workflows, including supplier verification, invoice processing, and compliance documentation, significantly reducing processing times while maintaining high accuracy (Tsang et al. 2024). Furthermore, predictive compliance monitoring systems facilitate continuous, automated verification of regulatory requirements (Rinderle-Ma, Winter, and Benzin 2023).

Prepare informational or reference materials.

Automation and robotics integration in logistics and intralogistics significantly impacts the task of preparing informational or reference materials by transitioning from manual documentation to digitally-managed, real-time information systems. Advanced warehouse management systems and Digital Twins now automate data aggregation from sensor data, eliminating manual transcription tasks while enhancing operational transparency (Li et al. 2023). Consequently, key performance improvements include enhanced system responsiveness and data synchronization through cloud-based architectures and supply chain integration. The implementation of Industry 4.0 technologies such as RFID and automated identification mechanisms further optimizes warehouse operations by enabling seamless data collection without manual intervention (Song and Wu 2024).

Prepare products for testing.

AI-based smart quality inspection systems integrate deep learning with automated monitoring to achieve defect detection accuracies exceeding 99 percent, significantly outperforming manual preparation and inspection methods (Sundaram and Zeid 2023).

Read work orders to determine material or setup requirements

Unified visual extraction frameworks (Wan et al. 2024) replace manual order reading with automated parsing of product specifications and catalog references. Deep reinforcement learning algorithms (Mahmoudinazlou et al. 2025) optimize order interpretation and dynamic task assignment based on these extracted material requirements. Automation impacts demonstrate measurable improvements: order reading accuracy increased to 98.5 percent, and error-free material requirement interpretation achieved at 99.8 percent (Wan et al. 2024). Integrated systems enable real-time order understanding and autonomous task routing.

Receive shipments.

The task of receiving shipments in logistics can be automated through computer vision for package sorting, achieving 97.2 percent accuracy (Han et al. 2020), and au-

tonomous mobile manipulation systems for unloading and depalletizing based on 3D perception (Aleotti et al. 2021).

Record production information.

Some examples of automating production information include the documentation of quantity, quality, type, weight, and test result data through computer vision and machine learning systems, with inspection accuracy reaching 99.8 percent (Sundaram and Zeid 2023). Integrated vision-based inspection systems with deep learning algorithms replace manual data recording by automatically detecting defects and autonomously capturing product specifications and material properties in real-time.

Record shipping information.

Robotic process automation tools replace manual data entry of weight, charges, and discrepancies, eliminating human error and accelerating order recording. Integration of AI-driven supply chain visibility platforms (Lam, Tang, and Wong 2024) enables real-time tracking of outgoing orders with automated status updates. Furthermore, the integration of RPA with machine learning automates critical administrative workflows, including supplier verification and invoice processing (Tsang et al. 2024).

Review work orders or schedules to determine operations or procedures.

Digital-twin–driven deep reinforcement learning can automate the real-time interpretation of work orders and schedules by jointly optimizing task assignment and AGV routing/path planning under dynamic shop-floor changes. In a tightly coupled architecture where the digital twin participates in both training and inference, policies adapt to schedule updates, layout reconfigurations, and disturbances, improving throughput and reducing completion and idle times compared with DRL-only or DT-assisted baselines (Lee, Kang, and Noh 2025).

Store records or related materials.

RFID-assisted AI identification systems (Song and Wu 2024) enable real-time inventory tracking and quality sensing, transforming manual record-keeping into autonomous, self-optimizing warehouse storage management. Automation impacts demonstrate measurable improvements, including space utilization increases of 60–80 percent and multi-criteria optimization incorporating turnover indices (Yerlikaya and Arıkan 2024).