DATA BRIEF

JOBS AT RISK — U.S. EMPLOYMENT IN THE NEW AGE OF AUTOMATION (PART I)

METHODOLOGICAL APPENDIX

METHODS AND DATA

Our method for estimating the fraction of U.S. employment that faces different levels of automation displacement risk assigns risk distributions to individual occupations based on data from the Occupational Information Network (O*NET)[1] and combines those risk assessments with employment level data from the U.S. Bureau of Labor Statistics' (BLS') May 2023 National Occupational Employment and Wage Estimates, which are produced as part of the BLS Occupational Employment and Wage Statistics (OEWS) program. This section will illustrate our approach at a high level using a stylized example.

O*NET DATA

Our measure of automation displacement risk relies on the "Degree of Automation" question in the O*NET Work Context module:

How automated is your current job?[2]

a. Not at all automated
b. Slightly automated
c. Moderately automated
d. Highly automated
e. Completely automated

Table A1: O*NET Degree of Automation and BLS OEWS Employment Data for 5 Hypothetical Occupations

O*NET Degree of Automation Response Shares (%)						
OCCUPATION	NOT AT ALL	SLIGHTLY	MODERATELY	HIGHLY	COMPLETELY	BLS OEWS EMPLOYMENT (MAY 2023)
	[1]	[2]	[3]	[4]	[5]	[6]
A	55.4	31.3	12.9	0.4	0	525,000
B	32.4	24.8	21.7	19.9	1.2	1,320,000
C	17.7	15.5	23.7	30.1	13	4,780,000
D	45.1	29.3	25.6	0	0	375,000
E	6.7	10.5	11.4	42.3	29.1	2,340,000

[1] O*NET is a widely used database that uses survey data to describe the characteristics of over 800 occupations in the United States. In this study, we use the "Degree of Automation" question in O*NET's Work Context module to estimate the distribution of automation displacement risk for each occupation in the O*NET data. The O*NET data is updated frequently; we use version 29.1, which was published in November 2024. The O*NET program is sponsored by the U.S. Department of Labor, Employment and Training Administration.
[2] O*NET surveys are completed by two types of respondents: occupational incumbents and occupational experts. The former are workers answering questions about their current job, and the latter are answering questions about an occupation of which they have detailed knowledge. The question stated here is the version presented to occupational incumbents. Occupational experts are asked "How automated is work in the occupation?" The same Likert scale response options are provided in both versions.

In publicly available O*NET data, the weighted[3] distribution of survey responses to the degree of automation question is provided. In our hypothetical example, these distributions are reported in columns 1 through 5 of Table A1. For example, this table indicates that 55.4% of O*NET respondents for the hypothetical occupation A selected “Not at all automated,” 31.3% selected “Slightly automated,” 12.9% selected “Moderately automated,” 0.4% selected “Highly automated,” and 0% selected “Completely automated.” For each occupation, we treat this distribution of responses as representing the probability that a given job in an occupation is currently not at all, slightly, moderately, highly, or completely automated. For instance, because 29.1% of respondents for occupation E selected “Completely automated,” we estimate that 29.1% of all existing jobs in occupation E exhibit this level of automation.

A central assumption of our approach to assessing near-term automation displacement risk is that a job’s current state of automation determines its probability of future displacement. More specifically, we assume that jobs that are currently “Not at all automated” face negligible risk of near-term displacement, whereas jobs that are currently slightly, moderately, highly, or completely automated face slight, moderate, high, or very high displacement risk, respectively. The basic justification for this assertion is that the technological, economic, and/or social leap required to replace labor via automation is directly connected to a job’s current level of automation. Positions that are already highly or completely automated are likely to only require a minor nudge for complete displacement, whereas jobs that are currently not at all automated would require huge technological advances and/or shifts in economic, regulatory, or social conditions to move toward complete automation.

BLS MAY 2023 NATIONAL OCCUPATIONAL EMPLOYMENT AND WAGE ESTIMATES

Having determined the probability that any given job in each occupation falls into each level of automation displacement risk, we turn to data on occupational employment to determine the aggregate number of jobs at each level of risk. For this purpose, we use occupational employment information from the May 2023 edition of the BLS National Occupational Employment and Wage Estimates, which is produced as part of the BLS OEWS program.

Merging these employment estimates with the O*NET data requires an occupational crosswalk, since although the two coding structures are very closely related and both based on the 2018 Standard Occupational Classification (SOC) system, the O*NET occupational codes use a slightly more detailed eight-digit structure. Fortunately, the BLS maintains a collection of crosswalks for translating between commonly used occupational coding structures, including one that maps O*NET codes to the six-digit SOC codes used in the BLS OEWS data. Column 6 of Table A1 illustrates what a merger of these data looks like for the five occupations in our hypothetical example.

[3] A detailed overview of the data collection practices and statistical methodology used in O*NET is available in the 2024 OMB clearance information for O*NET. In particular, Supporting Statement B: Statistical Methods provides a useful review of the O*NET survey sample and statistical methods employed.

Table A2: Classifying Current Employment for 5 Hypothetical Occupations by Automation Displacement Risk

Estimated Employment by Level of Automation Displacement Risk						
OCCUPATION	BLS OEWS EMPLOYMENT (MAY 2023)	NEGLIGIBLE	SLIGHT	MODERATE	HIGH	VERY HIGH
	[1]	[2]	[3]	[4]	[5]	[6]
A	525,000	290,850	164,325	67,725	2,100	0
B	1,320,000	427,680	327,360	286,440	262,680	15,840
C	4,780,000	846,060	740,900	1,132.860	1,438,780	621,400
D	375,000	169,125	109,975	96,000	0	0
E	2,340,000	156,780	245,700	266,760	989,820	680,940
TOTAL	9,340,000	1,890,495	1,588,160	1,849,785	2,693,380	1,318,180

Once these employment levels are known, it is straightforward to calculate the fraction of jobs in each occupation that fall into a given level of automation displacement risk. These calculations are shown in Table A2. For example, O*NET data from Table A1 indicates that 55.4% of jobs in occupation A are currently not at all automated; as such, we conclude that this fraction of employment in A (290,850 jobs) faces negligible near-term automation risk. Similarly, we estimate that 31.3% (164,325 jobs), 12.9% (67,725 jobs), 0.4% (2,100 jobs) and 0% (0 jobs) of employment in A faces slight, moderate, high, and very high near-term automation displacement risk, respectively. By completing similar calculations for occupations B through E and aggregating the values by level of automation risk, we conclude that these five hypothetical occupations collectively contain 1,890,495 jobs (20.2%) that face negligible risk of near-term automation displacement, 1,588,160 jobs (17%) that face slight risk, 1,849,785 jobs (19.8%) that face moderate risk, 2,693,380 jobs (28.8%) that face high risk, and 1,318,180 jobs (14.1%) that face very high risk.

Of the 831 occupations covered in the BLS May 2023 National Occupational Employment and Wage Estimates, the vast majority (742, or about 89%) can be directly matched to one or more O*NET occupations. In order to obtain estimates for the BLS occupations that cannot be directly matched to the O*NET data, the 742 occupations that were matched are aggregated to higher levels so that a match could be made.[4]

[4] For example, the occupation "entertainment and recreation managers, except gambling" (SOC code 11-9072) could not be matched at the full six-digit level, so it was instead matched at the five-digit level (i.e., 11-907) using an unweighted average of the O*NET degree of automation values matched to all codes in the 11-907 group that could be matched to the O*NET data at the six-digit level. In this specific case, there is only one other member of the 11-907 group (11-9071, "gambling managers"), and that occupation was matched to the O*NET data at the six-digit level. As such, "entertainment and recreation managers, except gambling" was assigned the same O*NET degree of automation information as "gambling managers." Of the 831 occupations in the BLS data, 742 were matched at the six-digit level, 51 were matched at the five-digit level, and 38 were matched at the four-digit level.

ESTIMATES FOR OCCUPATIONAL GROUPS AND INDUSTRIES

Having assigned O*NET degree of automation data to all 831 occupations in the BLS data, it is then straightforward to derive the distribution of automation displacement risk for any collection of occupations, including across all occupations and for specific subgroups of interest. In fact, the SOC coding structure makes aggregating from individual occupations to occupational groups very easy because the structure of each code is designed to group occupations in related groups at several different levels of aggregation. For the purposes of our analysis, we focus on the overall distribution of automation displacement risk across all occupations, as well as equivalent distributions for major occupational groups (identified by the first two digits in each SOC code). In all cases, we calculate the distribution of automation displacement risk for any occupational group as the employment-weighted mean of the automation displacement risk distributions of individual occupations within the group.

It is similarly straightforward to estimate the distribution of automation displacement risk for individual industries, because the BLS National Occupational Employment and Wage Estimates include separate tables in which industry employment is broken down by individual occupation. As such, once O*NET degree of automation data has been merged with the 831 individual occupations in the BLS data, that information can then be merged with industry employment information. Having done this, the automation displacement risk distribution of each industry is calculated as an employment-weighted mean of the automation displacement risk distributions of individual occupations within the industry. The industries we examine in this study are major North American Industry Classification System (NAICS) sectors, each of which is represented by a two-digit (or in some cases a collection of two-digit) NAICS codes.

BENEFITS AND LIMITATIONS

As with any other methodology, the approach discussed above has some key benefits and limitations. One central benefit is that we allow automation displacement risk to vary across jobs within the same occupation. Put another way, we do not classify each occupation in the data as having a fixed level of automation risk; instead, we estimate the probability that an individual job in each occupation falls into one of the five O*NET degree of automation categories, which results in a distribution of automation displacement risk for each occupation.

At first, this seems counterintuitive, since any two jobs in the same occupation are presumably similar. However, upon closer inspection, it seems highly likely that the adoption of automation technology (and, by extension, the presence of automation displacement risk) should vary substantially across jobs within the same occupation, for a number of reasons. For instance, automation tools often require significant capital investment, and whether or not such an investment is optimal (or even possible) will depend on factors such as firm size and profitability. Relatedly, the choice to adopt automation technology should also depend on the cost of labor-intensive alternatives. As such, firms with access to relatively cheap labor would likely have less incentive to adopt automation tools. For these reasons, we argue that instead of assigning a fixed level of automation risk to each individual occupation, it is more informative and realistic to estimate the distribution of automation risk within each occupation.

Another benefit of our approach is that, in contrast to many prior studies in this area, we do not impose any assumptions on the types of occupations and/or tasks that are automatable. For example, in their work on AI exposure, Kochhar et al. (2023) estimate the fraction of U.S. workers who have high exposure to AI in their jobs by first classifying all 41 items in the O*NET work activities module as having low, medium, or high AI exposure. Having made these classifications, occupations are assigned an AI exposure level based on the relative importance of work activities that have low, medium, and high exposure to AI. Although such an approach has its own positive attributes, one major issue is that the initial classification of work activities is inherently subjective and relies on the researchers accurately predicting the elements of an occupation that are most impacted by AI. As such, any misclassification at this step could significantly impact the final results. In contrast, our approach is based on assessing the distribution of current automation levels in each occupation and assuming that these automation levels are predictive of automation displacement. Although any assumption involves a measure of uncertainty, we feel that this conjecture (i.e., that a greater level of automation today implies greater risk of automation displacement in the future) is comparatively safe and plausible.

Despite these benefits, our approach does suffer from limitations, including several that are common to other work in this area. One especially notable concern is that the O*NET data is updated in waves; as a result, some degree-of-automation results are very recent, whereas others are years old. Because automation technology has evolved so rapidly in recent years, this limitation means that many of the occupations we examine have degree-of-automation information that does not capture these new developments and their impact on automation in the current workforce. Another issue with the O*NET data in general is that sample sizes are often small, which means that some estimates we rely on in our calculations are measured with low precision. Finally, although valuable, the O*NET degree-of-automation question can only capture a limited amount of information about automation in any given occupation. For example, what does it mean quantitatively for a job to be "highly" versus "completely" automated?

Because of the limitations, SHRM is pursuing a large-scale survey in 2025 to improve our understanding of automation and generative AI in the current U.S. workforce. By filling some of these knowledge gaps, we aim to refine and extend the initial results discussed in this data brief.