The Impact of Artificial Intelligence on the Labour Market in Lao People's Democratic Republic and Viet Nam

Francesco Carbonero – Fondazione CRC and University of Turin
Jeremy Davies - East Village Software Consultants
Ekkehard Ernst - ILO Research Department
Frank M. Fossen - University of Nevada, Reno, and IZA
Daniel Samaan - ILO Research Department
Alina Sorgner - John Cabot University, Rome, IfW Kiel, and IZA

August 9, 2021

Abstract

AI is transforming labor markets around the world. Existing research has focused its attention on advanced economies, but little is known for developing and least developed economies. Differences in the impact of AI on labor markets in different countries arise not only from differences in the prevalence of occupations, but also from the fact that occupations vary across countries in their composition of tasks. Estimates of the suitability for machine learning are provided by Brynjolfsson et al. (2019) for work activities collected in the O*NET database for the US. We propose a new methodology to translate these or similar measures of AI impacts to other countries at the level of work activities. To do so, we assess semantic similarities between textual descriptions of work activities from the O*NET database and workers' skills elicited in surveys available for developing and least developed countries, in particular the World Bank's STEP Skills Measurement Program. We illustrate our methodology for the case of Lao PDR, a least developed country, in comparison to the neighboring developing country Viet Nam. Our approach allows characterizing the extent to which jobs of individuals and occupations in these countries are subject to destructive digitalization, which puts workers at risk of being displaced, in contrast to transformative digitalization, which tends to benefit workers. We find that occupations in Lao PDR are less likely than in Viet Nam to be in machine terrain, where workers will have to adapt to transformed occupations and are at risk of being partially displaced.

Keywords: Artificial intelligence, machine learning, digitalization, labor, skills, Lao PDR, Viet Nam

JEL classification: J22, J23, O14, O33

1. Introduction

The impacts of digitalization and AI technologies on labor markets are multifaceted. Workers performing predominantly work activities that can be automated are at risk of being displaced by digital machines. However, occupations combining activities that cannot be automated with those that can are likely to be transformed. Workers in these occupations may benefit from working closely with new digital technologies rather than being displaced by machines.

Prior research has investigated the impact of new digital technologies on occupations primarily in the United States (Frey and Osborne, 2017; Brynjolfsson et al., 2018; Felten et al., 2019; Fossen and Sorgner, 2019a, 2019c) and in some cases in other developed countries (Arntz et al., 2016, 2017). These papers develop measures of the impact of digitalization on occupations in these countries and proceed by testing effects on wages and unemployment (Felten et al., 2019; Fossen and Sorgner, 2019c).

A first step towards a more detailed account of the impact of AI technologies on labour is to differentiate between different dimensions of labour market effects. In this respect, Fossen and Sorgner (2019a) suggest to classify new digital technologies in the United States into "destructive" and "transformative" technologies. This allows the authors to map occupations in the US according to the destructive impact of digitalization, which displaces human labour, and the transformative impact of digitalization, which reinstates labour (see Acemoglu and Restrepo, 2018, for a theoretical framework).

However, a significant gap in the literature is that little research has investigated the impacts of digitalization and robotization in developing countries (Carbonero et al., 2018). While emerging markets, including some of the world's poorest countries, are already using basic AI technologies, for instance, in smart farming, credit scoring and targeted advertising, advanced AI tech-

nologies are not yet widely adopted there. Yet, there exists a substantial potential for adoption of such technologies to leapfrog traditional development models (IFC, 2020). Use of digital technologies has accelerated in developing and even least developed countries, not least due to lockdown measures that governments implemented during the COVID-19 crisis. For example, the lockdowns have led many enterprises in the service sector to switch to digital processes in Lao PDR.[1] Similar efforts were made in the education sectors in which many activities were held online. These developments may have been a trigger for further digitalization efforts in the near future. Research applying occupation-level data for the United States to other countries typically points to a substantial risk for job destruction and an imminent job crisis, especially when analysing developing countries (Balliester and Elsheiki, 2018).

There are several issues that need to be considered when analyzing the impacts of digitalization in the context of developing countries. Applying the occupational digitalization scores developed for the United States in the context of developing countries might lead to significantly biased results, since the occupational tasks in developing countries might differ considerably from the occupational tasks of a similarly coded occupation in the United States (Arntz et al., 2017). In addition, reproducing the approaches by the aforementioned papers for other countries would be very costly, because these methods rely on extensive surveys of experts.

The goal of this paper is to develop a methodology that allows translating existing scores of AI impacts at the level of work activities to the contexts of other countries. Our approach allows comparing countries at vastly different levels of development – including low-income and least-developed economies – with respect to the destructive and transformative impacts of digitalization on their labor markets. We illustrate the method using the case of Lao PDR and compare it

[1] https://www.la.undp.org/content/lao_pdr/en/home/blog/2020/how-covid19-could-accelerate-digitalization-in-laos.html

with neighboring Viet Nam, a developing country in the same region, which has transformed from one of the poorest countries in the 1980s into a lower middle-income country today.

Due to vast differences in the structure and organization of work across countries, one cannot expect the impact of digitalization to be similar, especially when comparing developed and developing countries. When transferring estimated digitalization impact scores from the United States to other countries, we argue that adjusting for the occupational structure in different countries is insufficient. This is because workers perform a different mix of work activities within the same occupation in different countries. For example, teaching is an important part of the occupation of craftspeople in Germany because they teach apprentices, whereas teaching crafts is performed by teachers in schools in other countries. Another example is farmers: A large share of a farmer's work in a developing country may be manual field labor, whereas a farmer's workday in the United States is filled to a larger extent with accounting work. Therefore, the impact of artificial intelligence (AI) on farmers may be different across countries. However, the potential impact of AI on specific work activities is driven by technology, not the organization of occupations, and therefore comparable across countries.

Therefore, we develop a methodology using the level of work activities that are subcategories of job tasks instead of the occupational level, which allows us to significantly improve the applicability of the digitalization measures developed for the United States in countries other than the United States. To demonstrate the method, among the digitalization measures available, we choose the suitability of work activities for machine learning as reported by Brynjolfsson et al. (2018). We propose to assess the semantic similarities between the textual descriptions of work activities in the database of occupations in the US, O*NET, with skills elicited in surveys available for countries of interest. For illustration, we use the STEP ("Skills Towards Employa-

bility and Productivity") survey provided by the World Bank and the PIAAC ("Programme for the International Assessment of Adult Competencies") survey provided by the OECD. Using the textual similarities as weights, we translate the digitalization measures to the level of skills and then to the level of individual jobs in STEP. At this level, our approach allows us to characterize the impact of new digital technologies on the specific job of each individual worker in the sample, which opens up opportunities for a more detailed analysis of the impacts of digitalization on different types of workers than offered by more aggregate approaches. For example, we can study heterogeneous impacts by age, gender, and other individual-level characteristics in a given country, in our case Lao PDR and Viet Nam. We also aggregate the digitalization measures to the level of occupations in the chosen country, using the skills used by workers in the sample in each occupation. This results in digitalization impact scores for occupations in Lao PDR and Viet Nam that can be compared to other countries.

The picture that emerges from our approach is insightful and shows that the impact of AI on individual workers is polarized in Viet Nam, with most respondents moderately affected, but a significant number of workers at high risk of being displaced by digital technologies; in Lao PDR, the impact is more evenly distributed. The most common occupation reported by STEP respondents in Lao PDR, subsistence crop farming, has a comparably low suitability for machine learning presumably due to the non-routine manual tasks performed. The most common occupations in Viet Nam are more suitable for machine learning, in particular the occupations of shop salespersons, but also of agricultural laborers (according to the tasks they perform in Viet Nam) and garment and related trades workers. At the same time, shop salespersons perform a relatively large variety of tasks in Viet Nam, some of which cannot be automated; this makes it likely that this occupation will be transformed rather than completely automated. It should be noted that our

approach only makes an assessment regarding the impact of digitalization on jobs, not about the overall risk of automation. Mechanization, for instance, might be a more direct force for change in the context of occupations in subsistence crop farming in Lao PDR.

We also compare our AI impact scores to scores from a naive approach that transfers the scores from the United States to Lao PDR and Viet Nam at the level of occupations. In comparison to our proposed method based on work activities, the naive approach of using the impact scores for occupations in different countries seems to produce too much noise to derive meaningful insights.

2. Labor Market Structure and Trends in Lao PDR

To illustrate our method, we focus on Lao PDR, a least developed country, and compare the results to its neighboring country Viet Nam, a developing country. This section provides background information on the labor market in the focal country Lao PDR. According to information provided by the last Labour Force Survey (2017)[2], the working-age population of Lao PDR is made of 4.8 million people in a population of 6.9 million. More than 2 million people are less than 15 years old, one third of whom are between 10 and 14 years old, implying that many young people will enter the working age population in the near future. However, the formal labor force participation rate is rather low (40.8%) because the majority of individuals do not work in the formal sector and are not actively seeking a formal job. Of the 2.8 million people out of the labor force, 55% are women, while, among the employed individuals, 54% are male workers. The low employment stock and the inequalities in working opportunities pose severe concerns on the economic and social conditions of a large fraction of the population in Lao PDR.

[2] https://www.ilo.org/surveyLib/index.php/catalog/2032

A high prevalence of low-quality jobs also challenges a sustainable and solid path for economic growth in the country. Of the 1.8 million people in formal employment, almost 40% work as unpaid contributing family members. This is particularly dramatic for the young population (see Table 1): more than 300,000 unpaid contributing family workers are between 15 and 34 years old. Younger persons are also vulnerable in their job search, given an unemployment rate that is twice as large as the one of the older population (18% against 9%). This is reflected in the extremely large share of workers below 34 years of age who are discouraged in their job search (72%).

Urban and rural areas in Lao PDR reveal substantial differences in their labor market structure. Although the stock of employment in urban areas is smaller, rural areas have an unemployment rate that is three percentage points higher than the national mean. Moreover, urban areas differ among them as well, with the capital province having both the highest participation (69.2%) and unemployment (8.6%) rate (2018 Household and Youth Survey).

Concerning the sectoral distribution of employment, the primary sector absorbs one third of formal jobs in Lao PDR. However, this share seems to decline as job creation increases in secondary and tertiary sectors. Especially, the service and public sectors employ the vast majority of workers with a formal job and contribute to 40% of the annual GDP (World Bank national accounts data).

As regards the occupational distribution (Table 2), about 40% of workers are employed in agricultural occupations, concentrated in rural areas. The rest of the labor force is made of a combination of cognitive and manual occupations. Service occupations, which reveal a wage below the mean according to the LFS data, are mainly represented by women. However, managers and professionals also show a higher share of female workers compared to the aggregate

share of women among employees (46%). As we show in the following sections, the skills used by women, likely those employed in service tasks, are also those with a higher risk of being subject to machine learning.

Workers with less than lower secondary education account for almost three quarters of employment. Looking at gender disparities, male workers are more likely to have completed higher levels of education than female workers. Urban workers are also better educated than their rural counterparts. In the rural areas, the share of persons with less than primary education is almost half of the overall employed population (46%).

All in all, the labor market in Lao PDR presents a variety of challenges that result from a combination of low levels of human capital among the employed population, structural transformation in the sectoral make-up of GDP and inequalities in the workforce in terms of labor participation of women and youth. Against this background, the occupational transformation triggered by an increasingly digitalized world of work will pose additional challenges for the economy and labor market in Lao PDR.

3. Data and Methodology

3.1. Choice of Digitalization Measures

Several measures of impacts of new digital technologies on occupations in the United States have been suggested by recent literature. To illustrate our method, among the available digitalization measures that we will briefly discuss below, we choose the suitability of work activities for machine learning (ML) provided by Brynjolfsson et al. (2018). The reason for this is that their measure is available at a very narrow level of work activities, while other available measures are usually available at the level of wider job tasks or occupations.

Brynjolfsson and Mitchell (2017) concentrate on the capabilities of machine learning as it "currently exists". Furthermore, the rubric only concentrates on technical feasibility, not on the economic, organizational, legal, cultural, and societal factors influencing ML adoption. In this context they identify the eight key criteria to distinguish tasks suitable for ML from tasks where ML is less likely to be successful. On this basis Brynjolfsson et al. (2018) create a rubric of 23 questions used to determine the degree to which a detailed work activity (DWA) as defined in the O*NET database (compiled by the US Department of Labor) is suitable for machine learning (SML). Based on a survey the authors evaluate the potential for applying machine learning to the 2,069 DWAs, 18,156 tasks, and 964 occupations in the O*NET database. The authors use Crowdflower, a Human Intelligence Task (HIT) crowdsourcing platform, where each DWA is scored by 7 to 10 respondents with knowledge in the area. The 23 questions ask the respondents to evaluate each DWA on the basis of the eight criteria. Brynjolfsson et al. (2018) then aggregate their scores from the detailed work activities level to the task level and further to the occupation level in the United States weighted by importance as recorded in O*NET. The result is an average SML score for each US occupation.

Since the SML scores reported by these authors focus on the possibility of automation of activities currently performed by human workers, the average SML of the work activities performed in an occupation can be interpreted as destructive digitalization putting workers at risk of being displaced (see also Fossen and Sorgner, 2019c). In contrast, the standard deviation of SML scores across work activities performed within an occupation reflects transformative digitalization, because occupations combining activities that can be automated with activities that cannot be automated are likely to be reorganized (Brynjolfsson et al., 2018) and transformed rather than to displace workers. Workers in these occupations are more likely to benefit from their close

interaction with new digital technologies than to lose their jobs. The SML scores have the advantage that they are first generated at the level of detailed work activities in O*NET. These detailed work activities resemble the skills elicited in surveys like STEP and PIAAC, which facilitates the translation of these scores to other countries.

Alternative currently available measures of new digital technologies could also be applied within our methodological framework, but some adaption would be necessary. A second option are the AI Occupational Impact scores provided by Felten et al. (2018), potentially as a measure of transformative digitalization (Fossen and Sorgner, 2019a, 2019c). These scores are constructed on the abilities level in O*NET. However, there are only 52 abilities in O*NET, much less than detailed work activities. Moreover, the textual descriptions of abilities in O*NET seem to be quite dissimilar to the textual descriptions of skills provided in STEP or PIAAC, reflecting different concepts underlying these measures and, therefore, making these scores less suitable for applying our approach to STEP or PIAAC surveys. Nevertheless, our approach would be very useful with individual-level surveys measuring individual workers' abilities.

A third option are the computerization probability scores provided by Frey and Osborne (2017) as a measure of destructive digitalization. However, these probability scores are only available at the occupation level, so one would have to break these down to the level of work activities, potentially implying additional imprecision.[3] Alternatively, one would have to resort to the simple approach of transferring the measure to other countries at the occupation level, which does not seem to be accurate, as argued above. Table 3 summarizes the first three measures.

[3] One way to do so could be to regress the computerization probabilities at the occupation level on the nine bottleneck skills from O*NET identified by Frey and Osborne (2017). One could then obtain scores at the level of these nine skills and transfer them to other countries with data available on these (or similar) skills. This would allow the prediction of computerization risk at the occupation level in countries where data on occupations linked to the bottleneck skills are available. See Arntz et al. (2016, 2017) for a similar approach.

A fourth option is provided by Webb (2020). He develops a measure of exposure of occupations to AI technology by matching descriptions of work tasks in O*NET to the text of patents using text similarity measures. According to the author, the measure reflects labor-displacing (i.e., destructive) technologies. This procedure produces AI exposure scores at the O*NET task level; however, the author currently only provides the data aggregated to the occupation level on his website. A fifth option arises from the current work of Fabrizio Colella and Rafael Lalive, who are also constructing an automation index for each occupation using the skill requirements for each occupation and the current status of automatability of each skill. There is no working paper or data available yet.

3.2. Motivation for Developing a New Method

The aim of our proposed methodology is to use digitalization scores at the work activity level estimated for the United States for other countries, assuming that the same work activity can in principle be automated similarly in other countries, but taking into account that the same occupation is composed of different work activities in different countries. This accounts for the main critique in cross-country applications of digitalization scores.

In the bubble graph in Figure 1, which is adopted from Fossen and Sorgner (2019a), each bubble represents an occupation in the United States. The aim is to create similar bubble diagrams for different countries to compare the digitalization impact. Therefore, we have to aggregate the chosen digitalization impact scores from the work activity level to the occupation level in each country. The main challenge here is that country-specific occupational databases containing detailed information on occupational structure, such as O*NET in the US, are rarely availa-

ble.[4] If data on the composition of occupations were provided for other countries, and if the work activities were sufficiently similar to those in O*NET, one could aggregate the digitalization scores to the occupation level in these countries. These scores would reflect the actual use of the work activities in the occupations in each country (and not in the US). However, this approach is not feasible due to lack of such occupational databases.

Several studies have therefore relied on correction procedures. In particular, Arntz et al. (2016, 2017) adjust occupation-level computerization risk calculated for the US occupations (Frey and Osborne, 2017) by regressing them on individual- and job-specific characteristics from PIAAC or other national surveys available in the US and the country of interest. Then they use the estimated coefficients to make predictions of computerization risk for individual jobs and occupations in other countries. While the correction procedure partly accounts for peculiarities of national labor markets, it has several drawbacks. First, before the regression can be estimated, the occupational codes used in O*NET (SOC) have to be translated to the occupational codes in PIAAC (ISCO) using a crosswalk, and the latter codes are only available at the imprecise 2-digit level. Arntz et al. (2016, 2017) use a multiple imputation method to deal with this issue. Second, the approach starts with digitalization scores at the occupation level, whereas we suggest starting with scores directly attributed to the much finer detailed work activities to enhance accuracy and precision. Third, predictions from a regression have a lower variance than the original data, which is likely to be reflected in the results.

Therefore, in this paper, we rely on the main advantage of previous cross-country adjustment methods, namely the use of individual-level survey data, but aim to overcome the drawbacks of prior approaches mentioned above. Specifically, we propose to use representative indi-

[4] There are a few exceptions: for instance, occupational databases comparable to O*NET are available for Germany and Italy.

vidual-level surveys such as STEP (for developing countries) or PIAAC (for OECD countries). We use survey information on individual-level skills to assess the impact of digital technologies on individual workers instead of translating the digitalization scores from the United States to other countries at the level of occupations.

The STEP skills measurement program is provided by the World Bank and PIAAC by the OECD. The goal of both surveys is to provide representative individual-level data on the skills of the workforce and the usage of these skills in the individuals' jobs that can be compared across countries. STEP is based on the adult population aged between 15 and 64 residing in urban municipalities in developing countries. While the focus of PIAAC is primarily on high-income developed countries, the STEP survey focuses on developing and transition economies. So far, STEP has been administered in two waves, in 2012 and 2013, in 13 countries, including Lao PDR and Viet Nam. STEP surveys provide detailed information on individuals' socio-demographic characteristics (e.g., age, gender, formal education level) and job characteristics. For the purpose of our study, we use information on cognitive (e.g., reading and writing proficiency) and job-related skills (e.g., interpersonal skills or physical demand of jobs) that are relevant for performing a respondent's main occupation.

From the individual-level survey data, we can observe which skills are used in different occupations in each country. The skills reported by respondents working in each occupation in a given country can be used to compute skill compositions of occupations, which can then be the basis for the aggregation of the digitalization scores to the occupation level in different countries. One of the main additional advantages of our method is that researchers can choose between different levels of analysis. For instance, they can choose the skill level and create modified bubble graphs where each bubble represents a skill rather than an occupation. The size of a bubble

then reflects the share of workers with a given skill in the country. One could calculate these figures directly from PIAAC or STEP. To create the bubble graphs on the skill level, aggregation to the occupation level is not necessary. One would still need to match the O*NET work activities (that have SML scores) to similar skills in PIAAC and STEP (to translate them to other countries). Detailed analysis for different categories of workers, for instance, by age or gender, are also possible with the new method.

3.3. Matching O*NET Elements to Skills in STEP or PIAAC

A major challenge is how to match the descriptions of work activities in O*NET, which have digitalization scores such as the SML scores, to skills in PIAAC or STEP. At the level of abilities, a manual approach seems infeasible. For example, there are 52 O*NET abilities and 32 skills in STEP, so a translation matrix would require determining 1,664 weighting scores. Furthermore, this approach would be entirely subjective, as one would need to make an educated guess in each cell of the matrix.

We discuss three possible approaches. First, the simple approach is to use the occupational codes, which are available in STEP at a 3-digit level of ISCO-08 for most of the countries in the sample.[5] However, there are several disadvantages. Most importantly, occupations in developing countries could have different demand for skills compared to occupations in the United States, as discussed above. Furthermore, the level of aggregation is high, since the original data are at the 6-digit level of SOC and using a crosswalk from SOC to ISCO-08 results in a loss of precision. Therefore, to advance the literature, we aim to go beyond this approach and take the analysis to the level of work activities.

[5] There are several exceptions, e.g. Ukraine has 2-digit occupational codes and the Philippines use ISCO-88.

A second approach could be to conduct our own expert survey, similar to the approach of Brynjolfsson et al. (2018) and Felten et al. (2019), to produce new digitalization scores instead of using the scores published by these authors. One could conduct a survey among experts asking, for example, how much each of the nine AI fields used by Felten et al. (2019) will affect each of the skills elicited in PIAAC or STEP. This approach would make it possible to relate the AI progress rates directly to the skills elicited in PIAAC or STEP skills. A disadvantage is that this approach requires a substantial amount of resources (e.g., conducting a survey or collecting expert judgments). Moreover, the approach would be limited to specific databases such as PIAAC or STEP but could not be applied to other data sources after the survey is completed. We suggest this approach as a possible avenue for further research, which requires a long-time perspective and significant resources.

In this paper, we suggest and illustrate a third approach. We directly assign the about 2000 detailed work activities in O*NET to the STEP skills creating a matrix of relatedness. The PIAAC survey could also be used instead of the STEP to target a different set of countries. To overcome the issue of manual assignment of similarity mentioned above, we apply semantic text similarity matching techniques.[6] The idea is to find the semantic similarities between the descriptions of O*NET work activities and the STEP skill measures using an algorithm. This allows us to multiply the digitalization scores available for work activities with the obtained similarity measures, and then sum the products up, to obtain the skill-related digitalization measure. As mentioned above, in case of STEP, the text descriptions of the skills are more similar to the work activities than the abilities in O*NET. Thus, the approach works better at the work activities level than the abilities level. The SML scores of Brynjolfsson et al. (2018) are available at the work activities level, so the application of the approach at the work activities level is straightforward.

[6] Webb (2020) also uses a similar idea of matching textual descriptions of patents to tasks.

The main advantages of this approach are the following. It is systematic rather than subjective. There is no need to conduct new surveys. The same method can be used with different data sources such as STEP and PIAAC.

The main disadvantage is that O*NET provides separate elements at the level of abilities, tasks, and work activities, while STEP and PIAAC measure is a mix of these, which we refer to as "skills". The consequence is that the descriptions of the O*NET elements are often quite different from the descriptions of the STEP and PIAAC skills. Therefore, a pure text mining and similarity matching techniques would not identify a sufficient number of overlaps between the text description of O*NET elements and the text description of skill measures in STEP. In particular, synonyms used in the different databases such as "understand" versus "comprehend", as well as homonyms such as form (paper form or shape) are challenges to text similarity measurement. Our solution is to use state-of-the art semantic similarity matching, as described further below.

An additional challenge of this approach is that while O*NET represents a fairly complete set of elements (tasks, work activities, or abilities), individual-level surveys might be incomplete. For instance, for cost reasons, the surveys might decide to focus only on a few most important skills and not asking questions about other, less frequent skills. The matching procedure might therefore result in an imprecise estimation of digitalization impacts. This is particularly relevant when aggregating the digitalization measures to the level of occupations.

In this paper, we pursue the third approach. Our goal is to provide a new methodology and to illustrate its application. We also compare the results from this approach to results obtained from the first, simple approach of remaining at the occupation level.

3.4. A New Method Based on Semantic Text Similarity Matching

In this section, we describe our approach in detailed steps. The first step involves processing the textual descriptions of the detailed work activities (DWAs) in O*NET and the descriptions of the skills used by employed STEP respondents in their main job. The latter are the questions from the STEP questionnaire from the block of questions that aim at assessing the skills of employed respondents (see Table A1 in the Appendix). We combine the textual descriptions to a single string vector. Then we preprocess the string data stored in this vector. This includes removal of accents, HTML related information, consecutive whitespaces, substitutions of various text characters (e.g., " - ", "," "."), text conversion to lowercase. In the next step, word (semantic) embeddings are created for both DWAs and STEP questions using the Sentence-BERT (SBERT) method (Reimers & Gurevych 2019).[7] The model we apply is provided by MS Marco, which is pre-trained with real user search queries using the Bing search engine (a corpus that consists of 8.8 million passages).

In the second step, we create a similarity matrix that contains Cosine measures of similarity[8] between all documents in the sample using the semantic word embeddings created in the previous step. These similarity measures account for semantic similarity between the textual descriptions of DWAs from O*NET and STEP questions aimed at measuring workers' skills. In addition, we create a second Cosine similarity matrix between broader categories of DWAs and

[7] SBERT method is a state-of-the-art method in Natural Language Processing (NLP). It performs significantly better than alternative methods, such as averaging over a sentence's individual word embeddings and BERT (Reimers and Gurevych, 2019). The method has been applied, for instance, in the context of patent application (Jansson and Navrozidis, 2020) and gender differences in Covid-19 discourse on online discussion platforms (Aggarwal et al., 2020).

[8] Cosine similarity measure can take values between -1 and 1, where 1 means that two vectors of word embeddings point in exactly the same direction, -1 means that the vectors point in opposite directions, and 0 means that the vectors are perpendicular. We normalize the cosine similarity measures to take values between 0 and 1, which allows us to use them as weights later when translating the SML scores from the level of DWAs to the level of STEP skills.

all STEP skills questions to add more information on the nature of each work activity.[9] Our final similarity measures are built as an unweighted average between the two similarity measures between STEP skills and DWAs and STEP skills and DWA's broader category.[10]

Third, we use our final similarity measures as weights to create SML scores at the level of STEP skills. We do so by calculating for each STEP skill a weighted average of the SML scores defined at the level of O*NET DWAs:

$$SML_{skill} = \sum_{activities} \left(similarity_{activity,skill}\ SML_{activity}\right) \Big/ \sum_{activities} \left(similarity_{activity,skill}\right)$$

Fourth, we merge the SML scores calculated at the level of STEP skills with the individual-level STEP surveys for Lao PDR and Viet Nam conducted in 2012. There are three types of questions in STEP that are used to measure the skills respondents use in their jobs: yes/no questions about whether a certain skill is relevant in one's job (e.g., if a job requires reading books); cards questions that measure on a 10-point Likert scale the extent to which a particular job characteristic is relevant for one's main job (e.g., the extent to which a job is physically demanding); and frequency questions that measure (on a 4- or 5-point Likert scale) the time that a person dedicates to a particular skill/task in his or her main job. In order to make the responses to the different types of questions comparable, we normalize them such that the responses can take values within an interval between 0 and 1. Now we use the normalized individual responses to create a score capturing the SML of the skills each individual uses in his or her job. More

[9] Consider the DWA "Prepare forms or applications." We improve similarity matching results by adding information that this DWA belongs to a broader category "Documenting/Recording Information" instead of, for instance, a broader category "Handling and Moving Objects."

[10] Consider the example of the final similarity scores for the STEP question "Do you (did you) read anything at this work, including very short notes or instructions that are only a few sentences long?" The highest similarity score (0.749) is obtained for the DWA "Receive information or instructions for performing work assignments" and the lowest similarity score (0.181) is obtained for the DWA "Drive passenger vehicles."

precisely, we create an SML score for each individual *i* averaged over the skills and weighted by the normalized individual responses to the questions on the usage of these skills:

$$SML_i = \sum_{skills} \left(usage_{i,skill}\ SML_{skill}\right) \Big/ \sum_{skills} \left(usage_{i,skill}\right)$$

This is our measure of labor-replacing (destructive) AI technology at the level of the individuals' jobs.

Fifth, we create mean SML scores at the occupation level and the within-occupation standard deviation of the SML scores. We follow the method by Brynjolfsson et al. (2018) as closely as possible. These authors start with SML scores for each DWA in O*NET, which they aggregate to a broader level of tasks, and then average them within occupation (they call this mSML). In addition, they calculate the standard deviation of SML across tasks within each occupation (sdmSML), both weighted by their importance in the occupation as provided in O*NET. Since detailed occupation databases like O*NET are unavailable for most countries, including Lao PDR and Viet Nam, we use the STEP survey to derive the task composition of occupations in these countries. To do so, we calculate the average of the usage of each skill, obtained from questions in STEP, over individuals *i* in each occupation *occ* in a country:

$$usage_{occ,skill} = \overline{usage_{\iota,skill}}_{occ_i=occ}$$

Then we create an SML score for each occupation as the average SML score over the skills, weighted by the average usage of the skills in the occupation:

$$mSML_{occ} = \sum_{skills} \left(usage_{occ,skill}\ SML_{skill}\right) \Big/ \sum_{skills} \left(usage_{occ,skill}\right)$$

Finally, we calculate the standard deviation of the SML scores across the skills in each occupation, weighted by the average usage of the skills in the occupation in the country:

$$sdmSML_{occ} = \sigma(SML_{skill})\ weighted\ by\ usage_{occ,skill}$$

4. Results

In this section, we present the measures of destructive and transformative digitalization estimated for Lao PDR and Viet Nam following our proposed methodology. Figure 2 shows the kernel density of the suitability for machine learning of the skills reported by individuals in the STEP survey (SML_i) in both countries. For Viet Nam, there is a bimodal distribution: Most respondents have a mix of skills that is moderately suitable for machine learning, which shows that these individuals are at moderate risk of being displaced by digital machines. However, a significant number also exhibits skills that are highly suitable for machine learning. This points toward the fact that the labor market in Viet Nam's is more polarized than in Lao PDR in terms of susceptibility of individual workers' jobs to labor-displacing machine learning technologies.

Our method also allows us to aggregate by individual characteristics such as gender or age. Figure 3 suggests that in both Lao PDR and Viet Nam, women use skills in their jobs that are somewhat more suitable for machine learning than men. An interesting observation is that labor market polarization in Viet Nam is not specific to male or female workers, but it is a rather general phenomenon.

Several observations emerge when the data are disaggregated by workers in different age cohorts. In Lao PDR, workers in the youngest age cohort (less than 25 years old) seem to use skills in their jobs that are least suitable for machine learning than older cohorts, as shown in Figure 4. This appears to be different in Viet Nam, where high SML scores are most concentrated among individuals between the ages of 25 and 35. In addition, younger workers in Viet Nam seem to be the main driver of the observed labor market polarization in the country. In contrast, workers in the oldest age cohort (55-65 years old) are only moderately affected by machine learning technologies in Viet Nam.

Next, we aggregate to the occupation level in Lao PDR and Viet Nam. This tilts the distribution more to higher SML scores in Viet Nam (Figure 5). The mass of individuals with more moderate SML scores we saw in Figure 2 seems to be concentrated in a few occupations, such that more of the mass of occupations is concentrated at higher SML scores. Aggregation at the occupation level enables us to not only estimate the mean SML score (mSML) of the skills used in an occupation, but also the standard deviation of the SML scores (sdmSML) of the skills used within an occupation.

As argued above, if the skills used in an occupation can be automated on average, workers are at risk of displacement, so mSML is a measure of destructive digitalization. However, if some skills used in an occupation can be automated whereas others cannot, resulting in a high sdmSML score, the occupation is likely to get transformed (Brynjolfsson et al., 2018) and workers may benefit from increased productivity. Thus, sdmSML is a measure of transformative digitalization. Figure 6 shows that the distribution of sdmSML is shifted toward higher scores in Viet Nam in comparison to Lao PDR, which reveals that many occupations in Viet Nam are more likely to be transformed or reorganized due to AI than occupations in Lao PDR.

To visualize the effects of machine learning technologies on occupations in both countries, we map how much occupations in Lao PDR and Viet Nam are affected by destructive (mSML) and transformative (sdmSML) digitalization. In Figure 7, we can depict each occupation in Lao PDR and in Viet Nam as a bubble on a two-dimensional pane. Each bubble represents one occupation, and the size of the bubbles reflects the relative number of workers in the occupation in each country (approximated by the number of respondents in the STEP survey). We interpret high mean SML scores (on the horizontal axis) of an occupation as potential risk of automation, i.e. potential displacement of workers. A high variance of the SML scores (on the vertical axis)

shows that an occupation will be transformed: Some activities of this occupation can be carried out by machines, while others will continue to be performed by humans.

Occupations in both countries are roughly aligned along the first diagonal, indicating a tendency for significant labour market transformation: Occupations in the northeast corner are characterized by high transformative and destructive digitalization technologies, which places them in "machine terrain" with pressure on workers to adapt, combined with a risk of partial displacement. On the contrary, occupations close to the southwest corner show low SML scores and a low standard deviation in SML scores. These occupations can be considered to be in "human terrain", with little expected impact from AI. Few occupations are present in the northwest corner, which represent "rising stars" occupations, with limited risk of destruction and high potential for transformation. Similarly, very few occupations are placed in the southeast corner of "collapsing occupations" with high risk of destruction and little potential for transformation involving human workers (see Fossen and Sorgner, 2019a, for the characterization of the four sectors).

Comparing the two countries, many occupations in Viet Nam that are important in terms of employment are more suitable for machine learning than many important occupations in Lao PDR and thus potentially subject to destructive digitalization. Many of these occupations in Viet Nam are subject to transformative digitalization at the same time and located in the upper right corner of the chart characterized as machine terrain. The most common occupations among the STEP respondents in Viet Nam, represented by the largest bubbles in the graph for Viet Nam, are (in this order) "garment and related trades workers", "shop salespersons", and "agricultural, forestry and fishery laborers". These three occupations have relatively high SML scores in Viet Nam due to the activities performed in these occupations there. Among these three occupations, the activities of shop salespersons are most suitable for machine learning on average. However,

at the same time, the activities performed within this occupation have the highest standard deviation of SML, which suggests that the occupation will be reorganized and human workers will still be needed in this occupation in the future to perform some of the activities. In contrast, in Lao PDR, by far the largest share of STEP respondents work as "subsistence crop farmers" (large bubble in the graph for Lao PDR). The suitability for machine learning is lower in comparison to the above-mentioned occupations due to the manual non-routine tasks performed. However, in Viet Nam, some occupations are also located in the lower left quadrant, for example, "domestic, hotel and office cleaners and helpers" and "waiters and bartenders", characterizing them as within 'human terrain' for the near future in this country with low levels of both destructive and transformative digitalization. This finding suggests a large polarization of occupations in Viet Nam, which may reflect that this country is currently undergoing a significant shift from traditional occupations to those affected by industrialization and digitalization. In contrast, employment in Lao PDR is still dominated to a large extent by agricultural occupations that lie somewhere in the middle on the scales of both transformative and destructive digitalization. In other words, workers in Lao PDR are currently facing a lower risk of digital transformation as the labour market has not yet experienced fully previous waves of automation.

In a nutshell, the polarization between 'machine terrain' and 'human terrain' occupations is clearly more pronounced in Viet Nam than in Lao PDR. At the same time, none of the two countries have many occupations that can be characterized as 'collapsing' occupations, which are strongly affected by labor-displacing AI, or that fall into the category of 'rising stars' occupations, which have low suitability for machine learning but at the same time a high potential for occupational transformation.

Finally, we compare the SML scores translated from the United States to Lao PDR and Viet Nam at the work activities and skills level following our approach to the SML scores simply transferred at the occupation level (naive approach). The naive approach requires applying a crosswalk between the SOC occupation codes provided for the SML scores by Brynjolfsson et al. (2018) and the ISCO-08 occupation codes available in the STEP survey. When we use this naive approach and transfer the SML scores from the United States to Lao PDR and Viet Nam at the occupation level (Figure 8), the maps provide no clear patterns and seem to reflect mostly random noise due to measurement error. Note that the largest bubble in Figure 7 for Lao PDR representing subsistence crop farmers is absent from Figure 8, because Brynjolfsson et al. (2018) did not include this occupation in their analysis for the United States. As our method is based on activities performed within the occupation, it has the advantage that occupations not included in the original analysis can be examined as well, including new or reorganized occupations.

5. Discussion and Concluding Remarks

We proposed a methodology that allows meaningfully assessing new digital technologies' impacts on jobs in different countries. So far, the assessment of AI impacts on labor markets in countries other than the US has been rather limited, particularly so in developing countries. While the implementation of AI technologies is still rather low in developing countries, basic AI technologies are already in use there, and substantial potential for adoption of more advanced AI technologies has been identified (IFC, 2020). Pronounced interest in enhancing the implementation rate of AI technologies in developing countries is further driven by the promise of these technologies to help leapfrog development. Hence, understanding the impacts of AI on labor markets in developing countries, including in least developed countries, is crucial, but is dependent on the availability of appropriate methods. Previous methods that we discussed in this paper

do not sufficiently account for the fact that occupations are organized in different ways and comprise different work activities across countries. This has been the main challenge to the study of impacts of digitalization on occupations in various countries.

The novel method we propose in this paper relies on the assessment of the suitability for machine learning of more than 2,000 detailed work activities that constitute occupations. These detailed work activities are reasonably universal activities that can be considered relevant in all labor markets including those in developing and least developed countries. This highly disaggregated level of analysis allows us to aggregate digitalization measures in different ways, thus, overcoming the main challenge described above. In a nutshell, our method assesses semantic similarities between textual descriptions of detailed work activities in the occupational database O*NET in the United States, for which digitalization measures are available, and skills elicited in household surveys available in a wide range of countries, such as STEP or PIAAC. This makes it possible to translate measures of digitalization to other countries at the level of work activities and to compare the impact of digitalization across countries and for various groups of individual workers within countries. This method builds on and advances prior approaches such as that suggested by Arntz et al. (2016, 2017), which starts from occupation-level digitalization scores instead of detailed work activities and relies on a crosswalk to 2-digit-level occupational scores.

We illustrate our approach using the suitability of work activities for machine learning (SML) provided by Brynjolfsson et al. (2018) as the digitalization measure, STEP as the survey of individual skills used at work, and the country cases of Lao PDR, a least developed country, and its neighbor Viet Nam, a developing country. Our methodology allows calculating digitalization scores at the level of individuals rather than at the level of occupations, and it provides less noisy and more insightful results than the naive approach when digitalization measures are trans-

lated to other countries at the occupation level. While suitability of work activities for machine learning reflects destructive (or labor-displacing) AI technology, we also calculate the within-occupational variation of this measure to account for transformative effects of AI technology or the extent to which an occupation can be reorganized rather than displaced.

The main insights from our analysis can be summarized as follows. First, we find that a larger share of individuals and occupations in Viet Nam are exposed to labor-displacing machine learning technologies than in Lao PDR. This observation might reflect the differences in skill use between the two countries but also the fact that Viet Nam has already seen a larger transformation of its labor market through previous waves of mechanization. A significant share of workers in Lao PDR are employed in subsistence farming where the immediate implementation of AI technologies is challenging given the current state of technology and human capital in the country. This reduces the threat of rising unemployment, but at the same time casts doubt on the feasibility of leapfrogging the current development path by means of AI technologies in Lao PDR.

Second, the labor market in Viet Nam is pronouncedly more polarized with respect to the impacts of AI on individual workers, as compared to the labor market in Lao PDR. Both countries have a high share of workers in occupations that are characterized by high suitability of work activities to machine learning technologies, and, at the same time, have a high potential for re-organization of tasks within occupations. However, in Viet Nam there are few relatively highly populated occupations – predominantly in services – that mainly consist of work activities that are not very suitable for machine learning technologies. While these occupations can be considered as safe in terms of labor-displacing effects of AI on them, there are not many opportunities for workers employed in these occupations to improve their productivity by means of AI.

Third, the results of gender-disaggregated analysis indicate that in both countries female workers are slightly more affected by labor-displacing AI technologies than their male counterparts. This is in line with previous research on the impacts of digital technologies on women in the context of developing countries (e.g., Sorgner, 2019). We further show that labor market polarization in Viet Nam does not seem to be driven by male or female workers, but that it is a rather general phenomenon in this country.

Fourth, several important insights emerge from our analysis disaggregated by workers in different age cohorts. We find substantial differences in both countries regarding the impact of AI technologies on younger workers. In Lao PDR, younger workers appear to be least affected by suitability of their work activities for machine learning technologies, while in Viet Nam younger workers seem to be among the most affected by this type of AI technology. This suggests that there are substantial differences in skill use among young workers in both countries, which deserves a more in-depth analysis given that particularly in Lao PDR the share of young individuals in the population is substantial. In Viet Nam, the labor market polarization seems to be partly driven by the high susceptibility of young workers to AI technologies.

Our analysis is not without limitations. Some limitations can be attributed to the methodology, while others are due to the data used in our analysis. In terms of methodology, we were able to improve earlier methods by significantly disaggregating the level of analysis and breaking it down to the level of detailed work activities. Still, one may wonder in how far the detailed work activities are comparable across countries, given different stages of economic development. We argue that more than 2,000 detailed work activities is currently the most disaggregated level of analysis used in the literature, which represents an important advantage of our method. The highly disaggregated level of work activities makes them rather universal and applicable in various

contexts. Another challenge with our methodology is the application of similarity matching techniques with textual data. We rely on the state-of-the-art Natural Language Processing technique, namely SBERT, to create semantic word embeddings to be used later for finding similar textual descriptions of work activities and skills. Should more advanced methods become available in the future, the method can be adjusted accordingly.

There are several limitations in terms of data used in the analysis. First, surveys like STEP and PIAAC elicit a rather restricted number of skills, which might lead to imprecise results of similarity matching with work activities, as some of the latter might be relevant for one's job but corresponding information is missing in the survey. Therefore, household survey programs should ensure to include comprehensive information about skills and tasks that do not miss important areas. Second, for illustration purposes we used the measure of suitability of work activities to machine learning provided by Brynjolfsson et al. (2017), because currently it is the only digitalization measure that is available at this narrow level. If other measures, for instance, of other types of AI technologies will be developed in the future that are available at this level of analysis, they can easily be adopted with our methodology. In addition, future surveys should also attempt to distinguish between work activities, tasks, and abilities in a more systematic way, because some existing AI measures are available at the level of abilities (e.g., Felten et al., 2018), which were not measured in the STEP survey and therefore could not be analyzed with our method. Third, the STEP surveys for Lao PDR and Viet Nam are only available for the year 2012. It would be very useful to have similar surveys of adult's skills in developing countries that are more recent, representative and include a sufficient number of respondents to allow for a meaningful analysis of different categories of workers. In addition, the STEP survey was only conducted in urban areas of developing countries but given a strong urban-rural regional divide

in these countries, it would be desirable to have data that also includes respondents residing in rural areas.

Our proposed methodology opens avenues for future research by allowing the estimation of digitalization impact measures of choice for a wide range of different countries, both developing and developed countries. While our illustrative example focuses on SML scores, the STEP survey and the case of Lao PDR in comparison to its neighbor Viet Nam, other digitalization measures, other surveys such as PIAAC, and other countries should be investigated in the future. The full value of our approach will become visible when applying it to various countries, because the methodology allows using the same digitalization measures across countries, which makes the results comparable. This research will inform policymakers about challenges and opportunities that new digital technologies deliver to different labor markets in a more targeted and precise way than current research does, which focuses primarily on the United States. Comparing the impact of digitalization between developed and developing countries will allow adjusting economic development strategies in a timely manner. Future research will also be able to apply our methodology to regions within countries such as US states or regions within member states of the European Union as far as surveys with sufficient size are available. This research will show how regional imbalances will be influenced by digitalization and AI in the near future and allow policymakers to develop mitigating and enabling labor market policies such as targeted training programs.

References

Acemoglu, D. and P. Restrepo (2018): The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, 108(6), 1488-1542.

Aggarwal, J., E. Rabinovich, and S. Stevenson (2020): Exploration of Gender Differences in COVID-19 Discourse on Reddit. arXiv:2008.05713v1.

Arntz, M., T. Gregory, and U. Zierahn (2016): The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. OECD Social, Employment and Migration Working Papers, No. 189, Paris: OECD Publishing.

Arntz, M., T. Gregory, and U. Zierahn (2017): Revisiting the Risk of Automation. *Economics Letters*, 159, 157-160.

Balliester, T., and A. Elsheiki (2018): The Future of Work: A Literature Review, ILO Research Department Working Paper, No. 29, Geneva: International Labour Organisation.

Brynjolfsson, E. and T. Mitchell (2017): What Can Machine Learning Do? Workforce Implications. *Science*, 358(6370), 1530-34.

Brynjolfsson, E., T. Mitchell, and D. Rock (2018): What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *American Economic Association Papers and Proceedings*, 108, 43-47.

Carbonero, F., E. Ernst, and E. Weber (2018): Robots Worldwide: The Impact of Automation on Employment and Trade. ILO Research Department Working Paper, No. 36, Geneva: International Labour Organisation.

Felten, E.W., M. Raj, and R. Seamans (2018): A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *American Economic Association Papers and Proceedings*, 108, 54-57.

Felten, E.W., M. Raj, and R. Seamans (2019): The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization. Working Paper available at SSRN: http://dx.doi.org/10.2139/ssrn.3368605.

Fossen, F.M. and A. Sorgner (2019a): Mapping the Future of Occupations: Transformative and Destructive Effects of New Digital Technologies on Jobs. *Foresight and STI Governance* 13(2), 10-18.

Fossen, F.M. and A. Sorgner (2019b): Digitalization of Work and Entry into Entrepreneurship. *Journal of Business Research*, https://doi.org/10.1016/j.jbusres.2019.09.019.

Fossen, F.M. and A. Sorgner (2019c): New Digital Technologies and Heterogeneous Employment and Wage Dynamics in the United States: Evidence from Individual-level Data. IZA Discussion Paper 12242, http://ftp.iza.org/dp12242.

Frey, C.B. and M.A. Osborne (2017): The Future of Employment: How Susceptible are Jobs to Computerization? *Technological Forecasting and Social Change*, 114: 254-280.

IFC (2020): Artificial Intelligence in Emerging Markets: Opportunities, Trends, and Emerging Business Models. https://www.ifc.org/wps/wcm/connect/95a40480-27b5-4b99-8b4c-7768ae6a53a2/AI-Report_Web.pdf?MOD=AJPERES&CVID=nhLrsRc

Jansson, H. and J. Navrozidis (2020): Using Natural Language Processing to Identify Similar Patent Documents. https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9008699&fileOId=9026407

Reimers, N. and Gurevych, I. (2019): Sentence-bert: Sentence Embeddings Using Siamese bert-networks. ArXiv, abs/1908.10084.

Sorgner A. (2019): The Impact of New Digital Technologies on Gender Equality in Developing Countries. Inclusive and Sustainable Industrial Development Working Paper Series, #20/2019.

Webb, M. (2020): The Impact of Artificial Intelligence on the Labor Market. Working Paper, Stanford University.

Tables

Table 1: Precarious Labour Market Status in Lao PDR by age

Age bands	Contributing family workers	Discouraged job-seekers
15-24	137.134	39.037
25-34	167.201	37.285
35-44	157.727	18.416
45-54	119.728	8.134
55-64	73.869	1.941
65+	17.495	282
Total	673.154	105.094

Source: Labour Force Survey 2017.

Table 2: Occupational Structure in Lao PDR

Occupations ISCO 1-digit	Employees	Share over the total	Share of women
Armed forces	53.366	3%	14%
Managers	189.436	11%	59%
Professionals	138.341	8%	55%
Technicians and associate professionals	46.110	3%	36%
Clerical support workers	33.574	2%	38%
Service and sales workers	180.285	10%	64%
Skilled agricultural, forestry and fishers	696.720	40%	47%
Craft and related trades workers	198.077	11%	41%
Plant and machine operators, and assemblers	77.375	4%	17%
Elementary occupations	135.366	8%	34%
Total	1.748.650	100%	46%

Source: Labour Force Survey 2017.

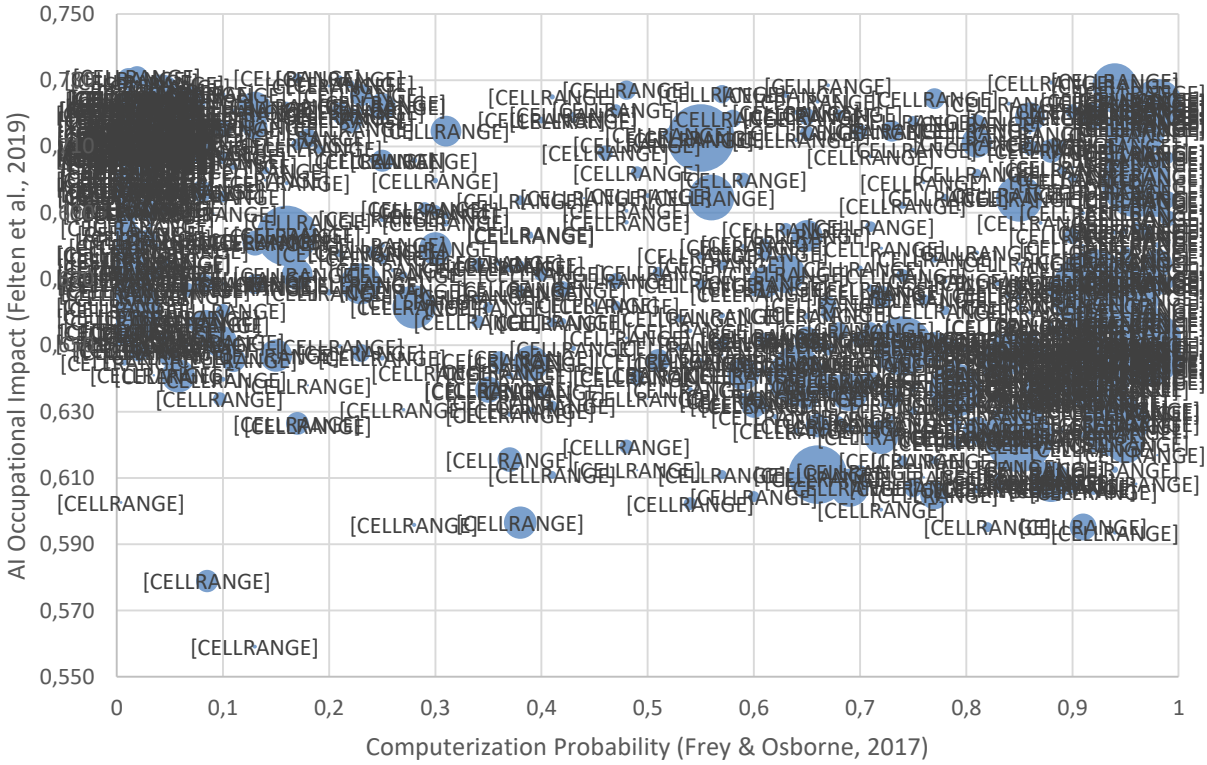
Table 3: Measures of Impacts of Digitalization on Occupations

	Computerization Probability	AI Occupational Impact	Suitability for Machine Learning
Short	CP	AIOI	SML. Within-occupation standard deviation: sdmSML.
Source	Frey and Osborne (2017)	Felten et al. (2019)	Brynjolfsson et al. (2018)
Time reference	Next 10-20 years (viewed from 2013)	2010-2015	Near future (viewed from 2018)
Focus	Digital automation	Artificial intelligence (AI)	Machine learning
Measurement	Experts' predictions for 71 occupations to obtain training dataset, then classification of 702 occupations using machine learning techniques.	AI progress measured by the Electronic Frontier Foundation mapped to 52 job requirements from O*NET and then aggregated to occupation level.	Scoring of 2069 detailed work activities from O*NET through a crowdsourcing platform, then aggregation to the occupational level.
We use the measure to operationalize	Destructive digitalization	Transformative digitalization	SML: Destructive digitalization sdmSML: Transformative digitalization

Notes: Overview of four measures of digitalization impact used in this analysis. Source: Fossen and Sorgner (2019c).

Figures

Figure 1: Computerization Probabilities of Occupations and AI Occ. Impact Scores in the US



Notes: Each bubble represents an occupation. The size of the bubbles reflects total US employment in the occupations (Bureau of Labor Statistics, 2018). Fossen and Sorgner (2019a) provide a similar graph, but they use scores in advances in AI provided by Felten et al. (2018), whereas this graph uses the updated AI Occupational Impact scores from Felten et al. (2019).

Figure 2: Suitability for Machine Learning of Indiv. Jobs in Lao PDR (Left) and Viet Nam (Right)

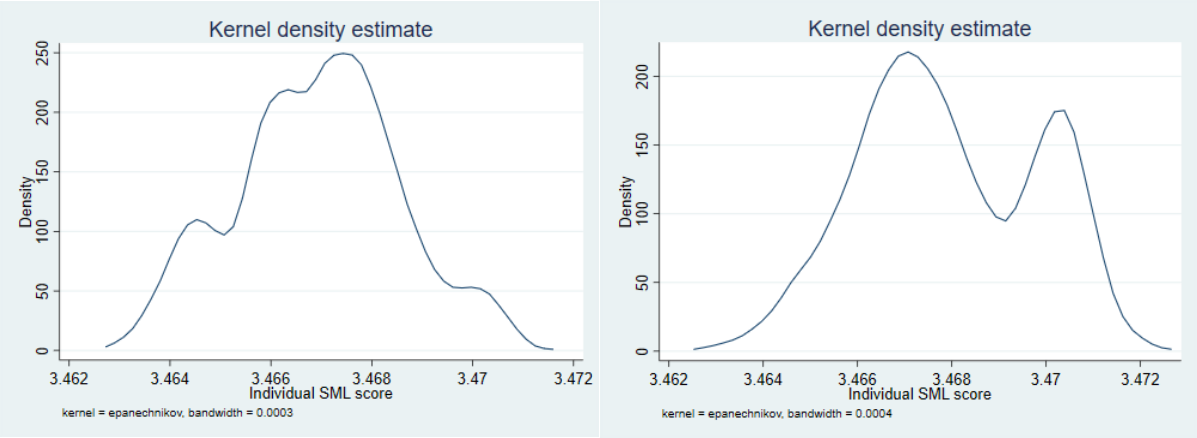


Figure 3: SML of Individual Jobs in Lao PDR (Left) and Viet Nam (Right) by Gender

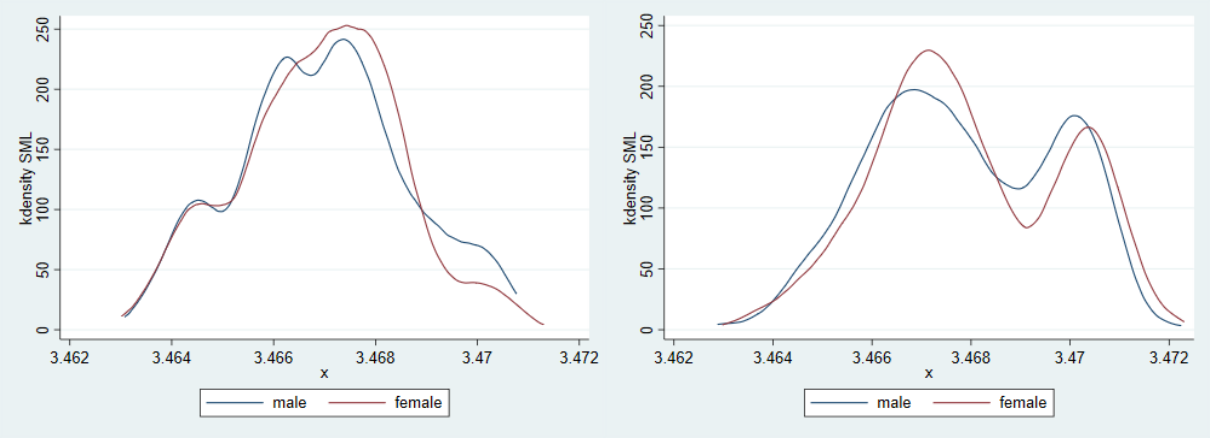


Figure 4: SML of Individual Jobs in Lao PDR (Left) and Viet Nam (Right) by Age

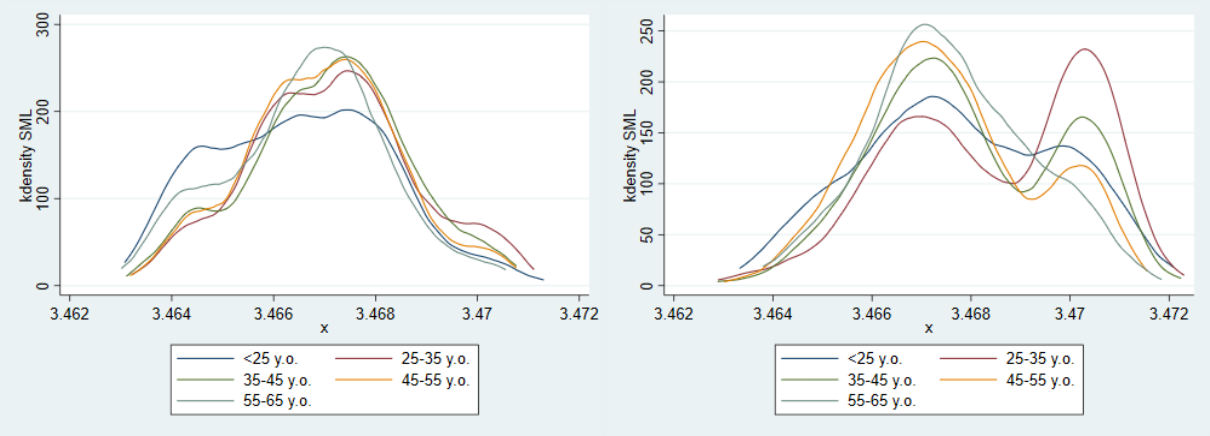


Figure 5: SML of Occupations in Lao PDR (Left) and Viet Nam (Right)

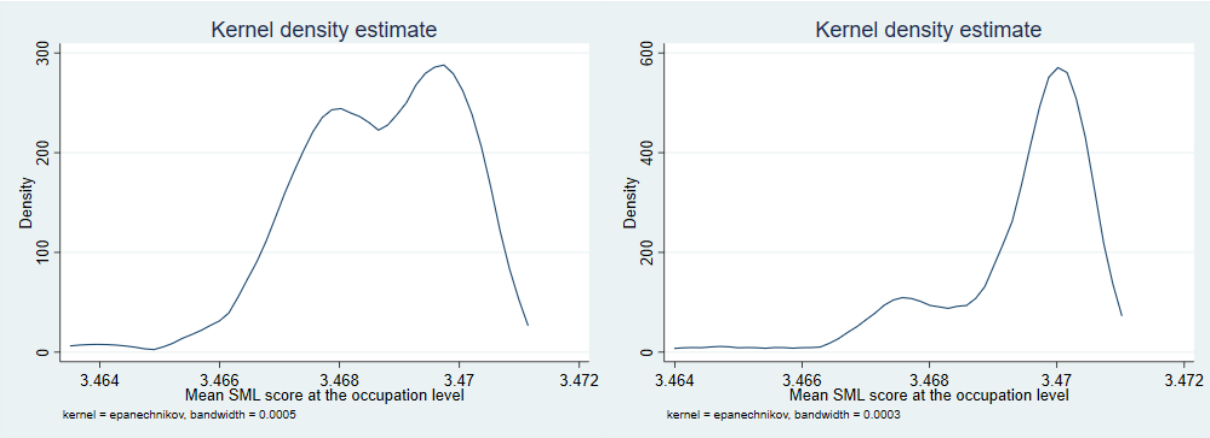


Figure 6: Within-occ. Standard Deviation of SML in Lao PDR (Left) and Viet Nam (Right)

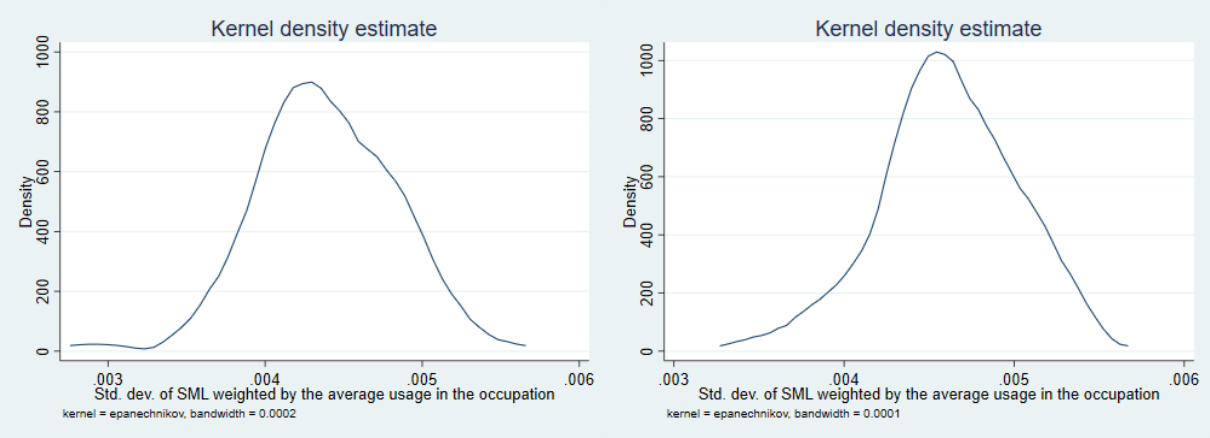
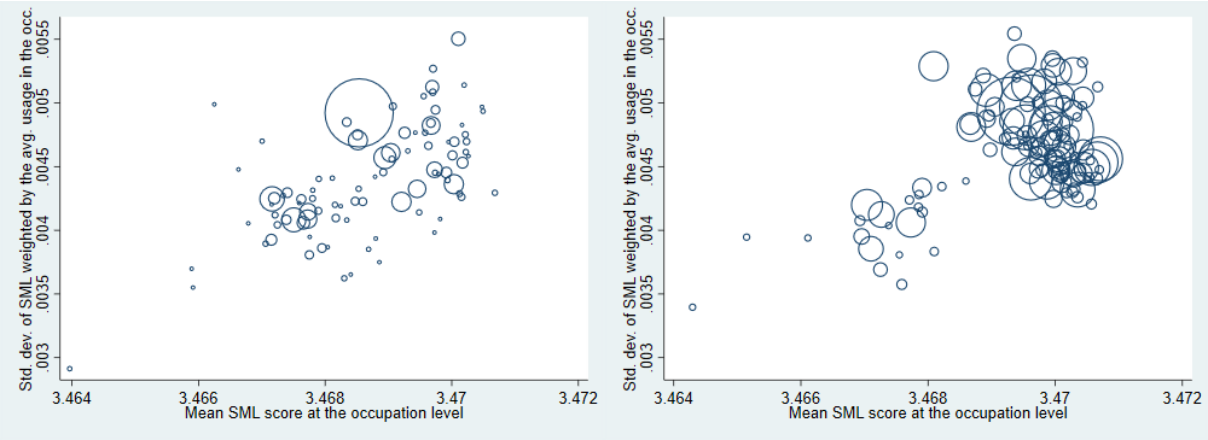
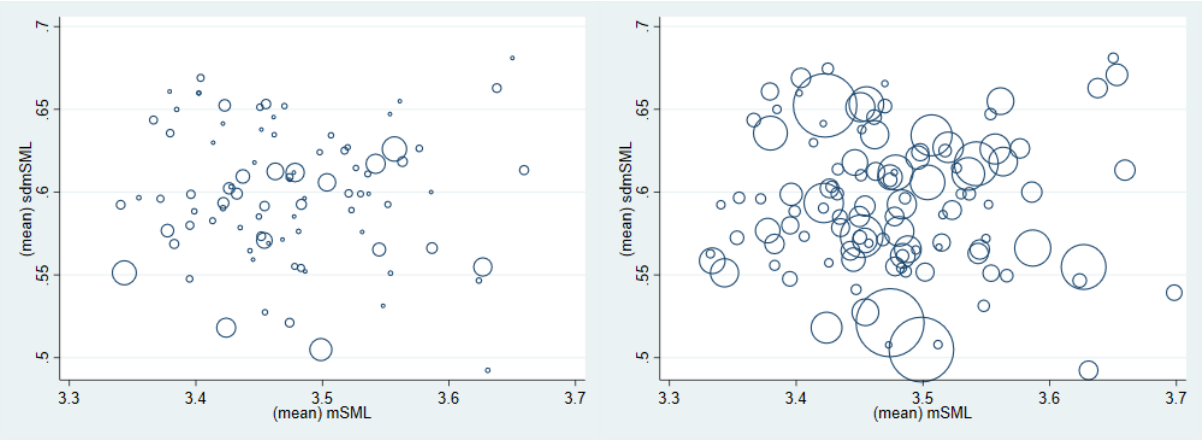


Figure 7: Destructive and Transformative Digitalization in Lao PDR (Left) and Viet Nam (Right)



Notes: Each dot represents an occupation in Lao PDR (Left) and Viet Nam (Right). mSML denotes the mean suitability for machine learning of skills used in an occupation and is a measure for destructive digitalization. sdmSML denotes the standard deviation of the SML of skills used within each occupation and is a measure of transformative digitalization. The red lines mark the means.

Figure 8: Transferring SML Scores from the US to Lao PDR (Left) and Viet Nam (Right) at the Occupation Level



Notes: Each dot represents an occupation in Lao PDR (Left) and Viet Nam (Right). mSML denotes the mean suitability for machine learning of skills used in an occupation and is a measure for destructive digitalization. sdmSML denotes the standard deviation of the SML of skills used within each occupation and is a measure of transformative digitalization. The red lines mark the means. The SML scores were translated from the United States to Lao PDR and Viet Nam at the occupation level (naive approach).

Appendix

Table A1: STEP Questions to Measure Workers' Skills

Question	STEP2012	STEP2013
Do you (did you) read anything at this work, including very short notes or instructions that are only a few sentences long?	1	1
As a regular part of this work, do you (did you) have to read forms?	1	1
As a regular part of this work, do you (did you) have to read BILLS OR FINANCIAL STATEMENTS?	1	1
As a regular part of this work, do you (did you) have to read INSTRUCTION MANUALS/ OPERATING MANUALS	1	1
As a regular part of this work, do you (did you) have to read BOOKS (OTHER THAN INSTRUCTION/OPERATING MANUALS)	0	1
As a regular part of this work, do you (did you) have to read REPORTS?	1	1
As a regular part of this work, do you (did you) have to read NEWSPAPERS OR MAGAZINES?	0	1
As a regular part of this work, do you (did you) have to read NEWSPAPERS, MAGAZINES, OR BOOKS?	1	0
As part of this work, do you (did you) fill out bills or forms?	1	1
Do you (did you) ever have to write anything (else) at work, including very short notes, lists, or instructions that are only a few sentences long?	1	1
As a normal part of this work, do you (did you) MEASURE OR ESTIMATE SIZES,WEIGHTS, DISTANCES, ETC	1	1
As a normal part of this work, do you (did you) CALCULATE PRICES OR COSTS	1	1
As a normal part of this work, do you (did you) PERFORM ANY OTHER MULTIPLICATION OR DIVISION	1	1
As a normal part of this work, do you (did you) USE OR CALCULATE FRACTIONS, DECIMALS OR PERCENTAGES	1	1
As a normal part of this work, do you (did you) USE MORE ADVANCED MATH, SUCH AS ALGEBRA, GEOMETRY, TRIGONOMETRY, ETC.	1	1
As part of this work, do you regularly have to lift or pull anything weighing at least 50 pounds [25 kilos]?	1	1
Using any number from 1 to 10 where 1 is not at all physically demanding (such as sitting at a desk answering a telephone) and 10 is extremely physically demanding (such as carrying heavy loads, construction worker, etc), what number would you use to rate how physically demanding your work is?	1	1
As part of this work, how frequently do you spend time co-operating or collaborating with co-workers?	0	1
As part of this work, do you (did you) have any contact with people other than co-workers, for example with customers, clients, students, or the public?	1	1
Using any number from 1 to 10, where 1 is little involvement or short routine involvements, and 10 means much of the work involves meeting or interacting for at least 10-15 minutes at a time with a customer, client, student or the public, what number would you use to rate this work?	1	1
As part of this work, do you drive a car, truck or three-wheeler?	1	1

As part of this work, do you (did you) repair/maintain electronic equipment? (cell phones, computers, printers, other electronic equipment...)	1	1
As part of this work, do you (did you) operate or work with any heavy machines or industrial equipment ? For example, machines/equipment in factories, construction sites, warehouses, repair shops or machine shops, industrial kitchens, some farming (tractors, harvesters, milking machine).	1	1
As part of this work, how often do you have to undertake tasks that require at least 30 minutes of thinking (examples: mechanic figuring out a car problem, budgeting for a business, teacher making a lesson plan, restaurant owner creating a new menu/dish for restaurant, dress maker designing a new dress)	1	1
As part of this work, do you (did you) have to make formal presentations to clients or colleagues to provide information or persuade them of your point of view?	1	1
As a normal part of this work do you direct and check the work of other workers (supervise)?	1	1
Still thinking of your work, how much freedom do you (did you) have to decide how to do your work in your own way, rather than following a fixed procedure or a supervisor's instructions? Use any number from 1 to 10 where 1 is no freedom and 10 is complete freedom.	1	1
How often does (did) this work involve carrying out short, repetitive tasks?	1	1
How often does (did) this work involve learning new things?	1	1
As part of this work do you (did you) regularly use A TELEPHONE, MOBILE PHONE, PAGER OR OTHER COMMUNICATION DEVICE?	1	1
As a part of your work do you (did you) use a computer?	1	1
As part of this work do you (did you) regularly use A BAR CODE READER?	1	0
Does (did) your work as [OCCUPATION] require the use of EMAIL	1	1
Does (did) your work as [OCCUPATION] require the use of SEARCHING FOR INFORMATION ON THE INTERNET	1	1
Does (did) your work as [OCCUPATION] require the use of DATA ENTRY	1	1
Does (did) your work as [OCCUPATION] require the use of WORD PROCESSING	1	1
Does (did) your work as [OCCUPATION] require the use of SPREADSHEETS (such as EXCEL)	1	1
Does (did) your work as [OCCUPATION] require the use of DATABASES (such as ACCESS)	1	1
Does (did) your work as [OCCUPATION] require the use of other software packages, designing websites or doing programming or managing networks?	1	1
Does (did) your work require the use of ADVANCED FUNCTIONS IN SPREADSHEETS SUCH AS MACROS AND COMPLEX EQUATIONS	1	1
Does (did) your work require the use of BOOK-KEEPING, ACCOUNTING OR FINANCIAL SOFTWARE	1	1
Does (did) your work require the use of PRESENTATION, GRAPHICS SOFTWARE (such as POWERPOINT)	1	1
Does (did) your work require the use of DESIGNING WEBSITES	1	1
Does (did) your work require the use of CAD SOFTWARE (computer aided design)	1	1
Does (did) your work require the use of STATISTICAL ANALYSIS OR OTHER ANALYSIS	1	1
Does (did) your work require the use of SOFTWARE PROGRAMMING	1	1
Does (did) your work require the use of MANAGING COMPUTER NETWORKS	1	1

Note: 1 indicates that a question is included and 0 that it is not included in a survey wave.