Unraveling Digital Automation in Emerging Economies:

A Problem Demarcation for Social and Economic Policymakers

Delft University of Technology

Master Thesis

Pascal Kampert

Unraveling Digital Automation in Emerging Economies:

A Problem Demarcation for Social and Economic Policymakers

by

Pascal Kampert

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday September 19th, 2023 at 11:00 AM.

Thesis Chair:	Prof. Jill Slinger, TU Delft
First Supervisor:	Dr. Enno Schröder, TU Delft
Second Supervisor:	Dr. Daniel Samaan, International Labor Organization
Project Duration:	March, 2023 - September, 2023
Faculty:	Faculty for Technology, Policy and Management, TU Delft

Cover:	Andrea De Santis, available for free use under the Unsplash License
Style:	TU Delft Report Style, with modifications by Daan Zwaneveld

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Executive Summary

Digisation and the continuous advancement of Artificial Intelligence (AI) incites projections about automation altering occupations over many sectors and countries. With increasing availability of digital infrastructure, many parts of emerging economies will open their doors and become subject to digitisation, robots and AI simultaneously (Messina 2016, Das and Hilgenstock 2018), which for advanced economies have appeared slower, with less established power structures and at different times. With high uncertainty around and little literature on the subject, the societal implications of automation in emerging economies are ambiguous, with potentials of automation boosting economies and leveling the playing field of development one one hand, but also leading to exploitative dynamics, in which companies with intellectual property rights and know-how on automation technologies inflict harm on labour markets, economies and thus many societal layers of emerging economies. While a multitude of indicators on job transformation through digitisation, robots or AI and their respective implications on employment emerged and were discussed within the literature for advanced economies, lack of abundant data causes ongoing uncertainty for emerging economies.

This thesis addresses the scientific gap of who is at risk of automation in emerging economies. The literature review entails an overview on the field of labour automation, the social effects of automation, and the (projected) case for emerging economies. The quantitative analysis entails the calculation of two automation indicators, Routine Task Intensity for automation through digitisation and robots, and Suitability for Machine Learning (SML) for automation through AI for emerging economies. SML and RTI are compared among occupations and socio-demographic groups as well as their relationship towards each other. To identify patterns to susceptible socio-demographic groups within the labour markets of emerging economies, a cluster analysis is performed on the most pertinent and yet orthogonal demographic parameters. The attained insights on socio-demographic clusters and their susceptibility towards automation are discussed to develop are broader picture on how automation affects emerging economies in order to educate policymakers and aid in rethinking policy approaches in response to changes in skill demands.

This thesis finds ground for concern that emerging economies will be subject to a not just broader but also likely more sudden wave of digital automation than advanced economies, potentially magnified through historic power imbalances and global market dynamics. Our results demonstrate that also for the task compositions of emerging economies, it can be expected that with the advent of AI a substantively larger share of tasks, occupations and socio-demographic groups is susceptible to automation as compared to automation through digitisation and robots. Workers with low and high education show higher automation susceptibilities than mid-educated workers, although through different technologies. The finding that also high-educated workers are susceptible motivates the conclusion that education systems might see their traditional societal role of enablers of social mobility endangered. Finally, this thesis finds that women are systematically more susceptible to AI-enabled automation and the observation that men are more susceptible to automation through digitisation and robots in advanced economies does not hold true for emerging economies. This thesis concludes by calling for particular consideration of emerging economies in the field of automation in labour markets, the merging of more granular data for emerging economies and an institutional model of automation.

Pascal Kampert
Geneva, September 2023

Contents

Introduction

The future effects of automation on labour markets in emerging economies are clouded by immense uncertainty. Contrary to advanced economies which currently prepare for automation through machine learning (ML), shortcomings in digital infrastructures of emerging economies still hold back automation through digitisation (Das and Hilgenstock 2018). With increasing availability of digital infrastructures, two connected automation technologies, digitisation and ML, reach labour markets at the same time and at a significantly riper stage than upon their emergence in advanced economies. In the frenzy of adopting general purpose technologies (GPTs) like artificial intelligence, labour markets repeatedly produced income inequality through job losses and wage suppression (Perez 2003), urging policymakers to anticipate and make well-informed decisions to mitigate undesirable developments.

The introduction of this thesis is divided in four sections: First, the background of this thesis is described. A short introduction into the research field of labour automation assists in the formulation of this thesis' research questions by unraveling the scientific knowledge gap this thesis contributes to close. Subsequently, the policy relevance of this knowledge gap is discussed, outlining the motivation of this thesis. The relevance of automation in academic and policy contexts is highlighted, as well as the contribution this thesis will make to both. Third, the research questions are stated and explained in synthesis, after which the fourth section presents an overview of the this thesis' general structure.

1.1. Background and Knowledge Gap

While earlier research exists, the field of (AI-enabled) automation risk in labour markets experienced a "big bang event" in 2013, as Carl Benedict Frey and Martin A. Osborne from the University of Oxford predicted 47 percent of US employees to be at risk of automation, in the first largely popular study in the field (Frey and Osborne 2013). Attention from outside the field of economics skyrocketed, sparking the public debate. Until now, the study remains the most cited paper in the field. Academic and commercial research subsequently grew in frequency and recognition. Scholars opposing the notion of widespread job losses in the near future either rejected methodological or semantic aspects of Frey and Osborne's study, sometimes leading to new estimation methods (Arntz, Gregory, and Zierahn 2017;Felten, Raj, and Seamans 2018; Brynjolfsson and Mitchell 2017), or to new perceptions of change: Many argued that the notion of AI directly causing job loss is challenged by the outlook of AI creating new jobs, changing the tasks constituting an occupation, which would lead AI to transforming jobs rather than automating them (Acemoglu and Restrepo 2018; Brynjolfsson and Mitchell 2017; Fossen and Sorgner 2019; Lane and Saint-Martin 2021). In short: While the field grew in numbers, the diversity of directions simultaneously increased, leading to a divided field in which many questions are yet to be answered.

Outside of academia, numerous popular authors attempted to either transparently communicate what is agreed upon and what is not, or to raise attention to the matter, warning of future job losses, and resulting societal shifts. Precht, R.D. (2022), accused labour economists of "futurising the past", arguing that current research collectively ignores the underlying threat of black swan events, rapidly changing labour markets with no opportunity to prepare, rendering predictions useless and creating necessity for more existential considerations in future labour market design. The rise of Chat-GPT, openAI's recently launched public interface to the GPT-3 language model, and the resulting discussion validates this position. In privately published books of the leading researchers, such as Brynjolfsson and McAfee (2014) and Frey (2019), these notions often are implicitly supported. Further emphasis can be layed on such criticisms as AI is currently not on the paths described by the initial predictions by Frey and Osborne (2013).

It is noteworthy that the current debate, whether sparked by Frey and Osborne (2013) or the release of ChatGPT, is not unique or the first of its kind within labour economics. In fact, the Frey and Osborne paper only does its reputation of being the foundation of AI-enabled automation justice, if the definition of AI is confined to that of machine learning (ML). Speaking broader, earlier available AI-systems, known as expert systems, were developed in the 1970s and found large-scale application in the 1980s (Waterman 1985; Leondes 2001). However, they simulated human decision-making in a – to today's standards – rather basic if-then logic, meaning a system was considered intelligent when it could in reaction to a circumstance (if) divert from its usual path and take accommodating actions (then) (Waterman 1985). Automation through such systems, or expert systems, will from now on be referred to as computerisation or digitisation. Computerisation has been subject of the debate on automation for twenty years (i.e. Autor, Levy, and Murnane 2003), even though it's effect on labour markets are dated back to the 1970s by some (Autor, Levy, and Murnane 2003; Goos and Manning 2007, Acemoglu and Autor 2011).

More sophisticated AI-models do not require the programming of if-then rules, but can deduce patterns themselves if provided enough high-quality data. Contrary to rule-based AI, ML-models enjoy a more powerful reputation when it comes to automation. The underlying logic is that while yes, prior waves of automation did not cause widespread unemployment, this time is different, because beyond certain bottlenecks like social tasks ”it is largely already technologically possible to automate almost any task” (Frey and Osborne 2013). Beyond its task-substituting nature AI also enjoys diverse economic benefits such as having strong first-mover advantages, being (close to) endlessly replicable at (close to) no marginal cost, and already being infrastructurally embedded anywhere where digital infrastructure is present, where AI-systems can be run on most PCs or even mobile phones (Ernst, Merola, and Samaan 2019).

For computerisation just as AI-enabled automation, task indices have been developed: Autor, Levy, and Murnane (2003) assessed the content of routine tasks, which can be conducted by a program or a robot, for occupations in the US labour market and found that occupations which characteristically had a large share of routine tasks (in other words: high Routine Task Intensity (RTI)) saw reductions in employment share and wage. These findings later on were replicated for many advanced economies (Goos and Manning 2007, Acemoglu and Autor 2011, Goos, Manning, and Salomons 2014). For AI, Frey and Osborne (2013) came up with a measure of automation risk, representing the “destructive” perspective, while Felten, Raj, and Seamans (2018) aim to measure transformative potentials in occupations, taking an epistemological counterpart. Brynjolfsson, Mitchell, and Rock (2018) conducted analysis on the replacement capabilities of ML on activities within occupation data taken from the US labour database O*NET, compiling average values on the suitability of ML (SML). These values have been assessed by mass-consultation of task- and AI experts respectively, which assessed the potential for 2018 ML to conduct individual tasks.

Unfortunately, what we know so far mainly focuses on the data-abundant advanced economies, especially the US, for which most quantitative analysis has been performed (Frey and Osborne 2013; Brynjolfsson and Mitchell 2017; Felten, Raj, and Seamans 2018; Acemoglu and Restrepo 2018; Fossen and Sorgner 2019). However, it is insufficient to transfer results of prior research directly onto emerging economies, as this would require to assume that occupations have the same task compositions as in OECD-countries, would neglect differences in educational make-up or worker skills and would assume similar patterns of technical diffusion. Yet, too few articles focus on the translation of these indicators to developing countries (Das and Hilgenstock 2018, Lewandowski et al. 2022; Carbonero, Davies, et al. 2023).

In fact, the effects of not digitisation but of AI on automation in emerging economies have barely been touched in the scientific field. Despite a thorough literature search only one paper discussing the topic was found (Carbonero, Davies, et al. 2023). Some researchers analysed computerisation potentials for emerging economies (Messina 2016, Das and Hilgenstock 2018, Lewandowski et al. 2022), showing that patterns of automation for advanced economies so far have not yet been reproduced on an economy-wide scale. As discussed by Das and Hilgenstock 2018, this is mainly attributed to a delay in development, which through the lack of digital infrastructure roll-out has prevented the automation of many tasks; investment into automation still significantly exceeds the low labour costs in emerging economies. Furthermore, no research so far presents an in-depth discussion of how an even broader wave of automation can be expected through the simultaneous introduction of digitisation and AI in emerging economies, both enabled through the same type of infrastructure.

1.2. Motivation and Contribution

As much of the knowledge on automation and its economical consequences has advanced on established economies, the knowledge vacuum regarding the impact to emerging economies persists. A consequence of this knowledge gap is particular uncertainty for policymakers. This is especially worrisome when considering that firstly, social security nets and labour regulations in developing countries oftentimes are weaker than in countries of the Global North, and secondly, AI-driven processes bear the potential to leapfrog in emerging economies as they require little investment and can be deployed as soon as they reach ripeness for market (Ernst, Merola, and Samaan 2019; International Finance Corporation 2020), as could be observed during the COVID-19 lockdowns (Homsombath 2020). Thirdly, informal sectors and potential for gig-economies are more prevalent is most developing economies, fostering employment precariousness (Anwar and Graham 2021; Stefano 2015). Overall, it is not unjustified to assume a harmful scenario imposing more grave consequences, to be managed with more scarce of resources, than in OECD countries.

A first step to unravel this knowledge gap is to analyse which range of occupations is susceptible to what type of automation in emerging economies. Accordingly, as part of this thesis the indicators of both Autor, Levy, and Murnane (2003) for routine-task automation and Brynjolfsson, Mitchell, and Rock (2018) for ML-automation are calculated in for emerging economies, which is this thesis' first contribution to the scientific body. Importantly, indices will be calculated in a way that allows for comparison of indices across emerging economies. This is made possible by use of the World Banks STEP-survey, which assessed task makeups of multiple thousands of individuals in multiple emerging economies. Moreover, while labour economists traditionally focus on occupations and qualifications when analysing labour markets – what this thesis also does – this thesis goes one step further to accommodate for characteristic traits of emerging economies: Lower strength and agility of educational and vocational systems in emerging economies render it more difficult to compensate for demand changes in labour markets. Being less able to reskill or upskill workers than advanced economies comparably magnifies the social consequences of employment changes, such as reductions in income or even loss of employment. Moreover, weaker social security systems further exacerbate undesired consequences for the working population. To not only account for social harms induced indirectly through changes in compositions of labour markets but to directly assess what implications automation can have on society this thesis considers automation directly through the social lens in its' second contribution: A socio-demographic cluster analysis reveals which social groups are particularly susceptible to either computerisation or AI-enabled automation. Anticipation of social effects of automation is key to be able to prevent increases of inequality between socio-demographic groups, induced by systemic labour market changes.

In doing so, this thesis contributes to the research on potential effects of digital automation on labour markets in emerging economies and provides insights that can help policymakers make informed decisions that promote economic growth, social stability, and equity.

1.3. Research Questions

In the following, this thesis' research questions, and their contribution to the existing literature are described. Concluding from the prior subsections, it becomes apparent that necessity for policy-relevant insights on automation potentials in developing economies emerged. The research questions are formed in recognition of the reservations in the popular literature, yet follow the current academic streams, hence building up on the body of literature for quantification and meaning of automation risks. The overarching question to be answered in this thesis is:

Who is at risk of automation in emerging economies,
and what are the implications for social policy-making?

The first research question addresses the scientific gap in quantification of automation risk for emerging economies. The second question builds up on the results of RQ1 with the aim to bridge a gap between quantified economic indicators and insights for policy design. For RQ2 it is analysed whether and which socio-demographic groups are particularly affected by automation, demarcating the societal threat imposed by digital labour automation. This thesis' discussion (Chapter 5) utilises the insights of the preceding research questions to form recommendations for further policy analysis.

RQ1: *Which occupations are likely to be affected by labour automation in emerging economies?*

RQ1 will be investigated through the calculation of RTI and SML for individuals and occupations in emerging economies. This will provide insights on which occupations and individuals are more likely to have a large share of their tasks automated, indicating either job or wage loss, or occupational transformation.

RQ2: *Which socio-demographic groups show the highest automation potentials for emerging economies?*

RQ2 lifts insights on individual traits to the meso-level by aggregating individuals to socio-demographic groups by means of a cluster analysis. The clusters are defined at hand of characteristics as age, sex, income, or status of employment, which then are related to either educational, economical or social policy measures.

The results of RQ1 and RQ2 are presented in the results section. The discussion section represents the main goal behind this thesis and serves to synthesise the quantitative results of RQ1 and RQ2 and further labour economic insights from the literature review into policy recommendations. These recommendations are not to be mistaken for a hands-on policy plan, but rather as recommendations of demarcating nature: Building up on prior insights, the aim is to identifying policy pathways which embody potentials to balance benefits of enhanced productivity and inhibit harms through unemployment or wage losses. The resulting policies recommendations hence are rather a set of promising options to be further explored in-depth. After all, the composition of economic factors is too distinct over different emerging economies to allow for one-size-fits-all solutions such as for instance in central bank policies. Every labour market policy is to be evaluated far beyond socio-demographic and macro-economic factors, incorporating further aspects like specifics of economic sectors, national labour law and cultural aspects. Yet, all specific labour market policies benefit of the initial general insights on who is at risk of automation, which this thesis contributes both through a labour market and through a social lense.

1.4. Overview

This thesis will first, in 2, provide a literature review on the labour market dynamics associated to AI-enabled automation. Within the literature review four questions are of core interest: How do economies adopt general purpose technologies (GPTs)? How does automation of tasks influence labour markets? What labour market consequences could be identified as a result of rule-based AI, and what consequences are projected for ML? What differences between emerging and developed labour markets play a role in automation? In synthesis, these four questions aim to provide the foundation for a story-line leading the reader from introductory dynamics of labour markets subject to technical change in the first two questions, and incrementally boiling down scope to this study: first to AI-enabled automation, then to AI-enabled automation in emerging markets in the latter two.

In the subsequent quantitative part, described in 3, two measures for AI-enabled automation potential will be calculated for emerging and developed economies: Routine Task Intensity (RTI) of occupations and individuals describing vulnerability to automation through rule-based AI (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011, Autor and Dorn 2013), and Suitability for Machine Learning (SML), describing the susceptibility towards ML-models (Brynjolfsson, Mitchell, and Rock 2018; Carbonero, Davies, et al. 2023). Subsequently, individual survey respondents from our datasets on emerging economies (data explained in: 3.1) will be clustered into socio-demographic groups using k-means clustering. The results of clustering of calculation of indicators will then be used to describe susceptibility to automation granulary in the results section.

In the discussion section, our results will be mapped against the insights obtained in 2. Here, one objective is to place our research in the context of the larger scholarly discourse, highlighting where there is overlap with and distinction to the existing body of literature for developed economies. Theoretical and practical implications will be discussed, focusing on how AI-enabled automation might affect labour markets of emerging markets under the assumption of high future diffusion of AI. The other goal is to identify which policy alternatives seem most relieving in light of the specific insights our socio-demographic analysis unveiled. We round of the discussion section by iterating through the limitations of our analysis.

Finally, we summarise and conclude this thesis' contribution to the scholarly discourse and, more importantly, its contribution to future policy intervention. Further out, the way for further research and policy analysis departing from this thesis' limited insights is described.

Literature Review

The body of academic literature initially was foraged systematically by classical means of a literature analysis in accordance with Creswell (2020). Through this, a significant part of available knowledge was discovered. Particular emphasis has been put on the identification of key concepts of automation in labour economics, and the respective developments in insights on such. In cooperation with researchers in the field from TU Delft and the International Labor Organization (ILO), central works have been backwards citation searched in Google Scholar and SCOPUS, unveiling the state-of-the art. Subsequently, the 47 individual papers retrieved were qualitatively assessed for subject relevance as well as academic appreciation in terms of journal and paper indices. Most importantly, Journal Impact Factor, h-index of authors, and paper citations with respect to their publishing date were considered, leaving us with 20 scientific articles which served as this thesis' foundation. But, more importantly, beyond these 20 articles, the process of writing this thesis was sidelined by regular research on individual topics which shape this thesis' line of research, incrementally providing more and more research with more than 90 academic sources supporting this thesis' train of thought.

This literature review aims to first provide this thesis with the foundation of how automation has been understood over time, how it has been analysed up to now and what relevance automation imposes on labour and consequently economic developments. The development of the field of automation economics is described chronologically. Accordingly, different forms of automation technologies, and how they influence labour markets will be discussed in order of their appearance. It is relevant to note that emphasis will be put on automation technologies which still are somewhat novel to emerging economies. While the potentials of older automation technologies such as the steam engine are irrefutable, they find no particular consideration in this literature review. After discussing what automation technologies are understood how by labour economists – research dominated by articles on western labour markets – and how what social implications automation holds, this literature review will discuss what is known about the case for emerging economies. The translation onto emerging economies will be shaped by the aim to derive policy-relevant insights, and will thus consider not only how automation has reached labour markets in emerging economies, but also what conditions specific to emerging economies surround policy arenas in emerging economies.

2.1. Central Constructs of Automation

Before this thesis dives into digital automation, it is relevant to note that the underlying concept of labour displacement is not specific to the emergence of artificial intelligence, or digital technologies in general. While the utopia of a work-free society has been discussed by philosophers dating back to ancient Greece, the scientific field of economics began considering replacement of labour through technology in the 1930s. This literature review will provide a chronological summary of unemployment through technological progress only within the academic sphere of labour and political economics.

2.1.1. Early Works

The question of whether and when labour finds an end through its economisation has first been raised 90 years ago by John Maynard Keynes (1933) in his essay ”Economic Possibilities for our Grandchildren”. According to Keynes (1933), this ”technological unemployment” will occur under the condition that economisation takes place faster than new uses for labour can be found. By providing an abstract conceptual vision, defining a sphere of labour which on one end expands into new facettes of labour and on the other end is reduced by labour being taken over through technological progress, Keynes (1933) brought the concept of “technological unemployment” into the scientific domain. This early investigation into the potential long-term effects of technological advancement laid the groundwork for further studies on the subject. Two decades after the second industrial revolution, or few years after the introduction of the five-day work week in Henry Ford’s factories facilitated by productivity spurts, full technological unemployment seemed like a logical consequence of growth and technological development. Yet, after the third industrial revolution and the emergence of industry 4.0, Keynes projection of large-scale technological unemployment could not hold true.

The Triple Revolution Report (Ferry 1964) elevated automation and its prospective impact onto the workforce into the spotlight of debates on public policy. The study, published as an open letter to the Johnson administration by the Center of the Study of Democratic Institutions, identified three upcoming civil revolutions: The weaponry revolution, the human rights revolution, and the cybernation revolution, which addresses the subject of technological unemployment and was given the highest emphasis out of the three revolutions. The core dynamic leading the cybernation revolution to be deemed as problematic was the uncertainty and inequality arising from unforeseen gains in productivity on one side, and the decline of necessity to include the public as workers on the other. Ultimately, the report encouraged President Johnson to counteract automation-induced societal frictions by promoting large-scale public works, affordable housing, public transportation, income redistribution, union representation for the unemployed, and restriction in the use of technology (Ferry 1964). While their implementation at the current time is beyond the scope of this thesis (interested readers are encouraged to read “The Economics of the Powerful” by Haring Douglas), it is important to note that these policy measures in the current day still enjoy relevance against the backdrop of future consequences of automation on labour markets (See: Chapter 6).

In 1975, Dutch economist Jan Tinbergen examined the factors impacting income distribution and assesses various policy solutions to reduce income inequality. To conceptually establish the ”race” between technology and education (also referred to as the supply of skills) is the book’s longest-lasting contribution. According to Tinbergen (1975), the interaction between technical improvements and the availability of skills in the job market causes economic disparity, building up on the empirical observation that those with low-education got systemically disadvantaged through technological progress. Later authors coined the term corresponding phenomenon as ”Skill-Biased Technological Change” (SBTC). As a result of technological advancement, demand for qualified labour rises, which may result in a more unequal income distribution. By increasing the supply of qualified employees who can compete in the labour market, increases in education and training, on the other hand, can contribute to a reduction of income disparity.

His book is the result of almost a decade of research and many individual studies performed by himself, in which he ultimately synthesised an early model on the trade-offs between high and low skill labour as well as technological capital in labour markets. High and low skill in this work are understood as levels of education, more particular work which requires higher education vs. work which does not. These two are understood as production factors and as such are imperfect substitutes; while high-skill workers can perform low-skill work and (in some cases) vice versa, their potential to substitute each other is constrained by the elasticity of substitution. The primary cause of changes in wage equality (or inequality) is attributed to technology, which is modelled as an exogenous third production factor. When putting technological progress aside, this elasticity between high and low skill labour is the key determinant for premia of high skill on income. This model, from now on referred to as the “canonical model” in accordance with Acemoglu and Autor (2011), has been widely adopted and thus well-established in the field of labour economics, and has been empirically supported. Numerous studies, namely Katz and Murphy (1992), Autor, Katz, and Krueger (1998) and Autor, Katz, and Kearney (2008), and Carneiro and L. (2009), demonstrate that it successfully accounts for key shifts in the distribution of incomes in the United States. Katz, Loveman, and Blanchflower (1995), Davis (1992), Murphy, Riddell, and Romer (1998), Card and Lemieux (2001), Fitzenberger and Kohn (2006), and Atkinson (2008) demonstrate that the model also succeeded in capturing significant cross-country variations across advanced nations.

2.1.2. Computerisation: Software and Robots as Enablers

Paradigm-changing in the economics of automation, Autor, Levy, and Murnane (2003) investigated what specific capabilities of computers had already automated tasks by the time of the paper’s creation. The paper serves as a response to a comprehensive body of literature providing empirical proof of a robust but unexplained correlation between labour requiring either higher or lower – but no mid-level education – and the adoption of computers in businesses. For this thesis, the paper provides two fundamental contributions.

The first is an epistemological clarification of the concept of work: First, Autor, Levy, and Murnane (2003) went beyond the approach of Tinbergen (1975), who considered labour as a classical production factor, and instead considered tasks individually. Considering individual tasks, Autor, Levy, and Murnane (2003) postulated that, contrary to humans, computers conduct each task as an independent process with a well-defined start, process and end. In other words, computers can understand tasks which can be programmed entirely. These tasks, coined as “routine tasks”, can be taken over by computers. For other “non-routine” tasks, computers are constrained by Polanyi’s Paradox (Polanyi 1966): “We know more than we can tell”. Ergo, humans capable of performing certain tasks, or possessing knowledge on them, cannot communicate what they know about these tasks in full extent. According to Autor, Levy, and Murnane (2003), the rules for “non-routine” tasks in consequence are not translatable to computer code. However, computers can augment or assist in performing these tasks, such as in calculations for problem solving or in communication. This contribution, diverting from the status quo distinction between high and low skill workers, challenges the canonical model and the SBTC-hypothesis and introduces Routine-biased technical change (RBTC) as a substitute theory.

Second, a simple model was proposed to explain the aforementioned observations by explaining the causality between the decline in cost for computer capital and the subsequent decline in demand for routine tasks: When the price of computational power decreases, occupations that initially require a high amounts of labour for routine tasks will make greater investment in them, resulting in a decline in the labour input required for substitutable routine tasks, and a rise in the input needed for supplementable nonroutine tasks. This dynamic leads to an increase in demand for either highly or lowly educated professionals, and thus comparative advantage over workers who perform routine tasks, which typically have mid-level educations. This model provides the theoretical foundation to the body of empirical evidence it responded to, and has since been taken up as the primary theory explaining the ”hollowing out of the middle” (See: Chapter 2.2.1)

In a chapter of the Handbook of Labour Economics, Acemoglu and Autor (2011) first built up on Autor, Levy, and Murnane (2003) and summarised trends identified within labour markets, devoting special attention to job polarisation; The observation that the distribution of skill levels shifts towards the respective ends, and the middle-skill occupations are in relative decline. Acemoglu and Autor (2011) follow RBTC, that “hollowing out” of the middle is primarily caused by computerisation, more specifically, by the decline of routine tasks, which characteristically were mostly allocated to middle skilled jobs. Tasks which are considered low-skill, such as driving a bus or painting a wall in comparison are less threatened, and thus remained in demand like high skill tasks performed by engineers and doctors did. The then canonical model was unable to explain these shifts, which among other reasons motivated the authors to develop and propose the tractable Ricardian model of the labour market to analyse labour market impacts of technological development under incorporation of task composition and other factors, such as offshoreability. Additionally, technological progress is endogenised in the Ricardian model. In doing so, they allow for understanding of speed and direction of technical change, which further on facilitates understanding incentives which drive technological adoption in companies.

2.1.3. The Arrival of Machine Learning in Labour Economic Theory

The single most famous paper on the topic, “The Future of Employment: How Susceptible Are Jobs to Computerisation?” written by Frey and Osborne (2013), imposed the next paradigm-change as the papers primary contribution expanded the scope of the analysis to include the potential impact of artificial intelligence and machine learning on both routine and non-routine tasks. The authors found that 47 percent of all US employment was at danger of automation in the future decades by using a novel method, building on a revisited task model which classifies tasks as susceptible or non-suceptible, to predict the probability of automation for 702 occupations.

On one hand, Frey and Osborne hereby challenged the belief that routine tasks are at the focus of automation, as their research shows that ML breakthroughs up until 2013 already proved the capability to also perform non-routine tasks. However, their findings should on the other hand not be understood as a full invalidation of Autor, Levy, and Murnane (2003) work; when applied to computerisation and automation by industrial machines Autor and Dorn’s insights remain uncontested. What Frey and Osborne (2013) primarily contributed was a challenge to the scope of prior discussion: Impacts of technological unemployment might be more extensive and far-reaching than previously imagined, because the substitutive potential of AI had not been not fully captured. In particular, the developing capability to automatically conduct a) non-routine cognitive tasks, which Autor, Levy, and Murnane (2003) and Frey and Osborne (2013) became increasingly feasible through progress in ML and Data Mining, and b) non-routine manual tasks, which by then could be performed through more sophisticated robots, which’s development had been spurred by advancements in sensing technologies as computer vision, ML-guided decision systems and their respective synthesis.

Frey and Osborne (2013) concludingly argue that close to all tasks technically could be automated and that the determining factor lies in where and when “engineering bottlenecks”, resembling skills for tasks which yet cannot be automated, can be overcome. Such engineering bottlenecks at the time, according to Frey and Osborne (2013), were skills such as human-level perception and manipulation (finger and manual dexterity, worked in cramped workspaces), social intelligence (negotiation, persuasion, social perceptiveness and caring for others) and creative intelligence (originality and fine arts). In consequence, the authors concluded that the trend of job polarisation will eventually be halted and subsequently be replaced by a trend of occupations slowly developing towards being comprised of social and creative tasks.

2.1.4. A Call to Nuance

Corresponding to the upstir the Frey and Osborne paper caused in scientific as well as popular literature, multiple answers, critiques and developments, emerged in the following.

While Autor (2015) does not directly disagree with Frey and Osborne's findings regarding the potential impact of AI and automation on employment, a response to the general warning undertone of the debate on automation in labour markets is provided. This response takes form in a more nuanced perspective on the relationship between automation, labour markets, and the future of work. Autor (2015) highlights the historical and theoretical context of workplace automation taking into account examples from the Industrial Revolution to the present day. He argues that contrary to Frey and Osborne (2013), who mainly focused on direct labour displacement, the primary effect to be observed in future still is that automation will likely supplement human labour in non-routine tasks as well as opening new possibilities for human-machine collaboration. This would lead to new forms of value creation and economic growth. More importantly, economic forces of supply and demand will play a critical role in shaping the impact of automation on wages and employment. To support his claim he examined empirical evidence on the labour market effects of technological change, highlighting the role of economic forces in shaping the distribution of wages and employment opportunities. Finally, he concludes that while AI-enabled automation undoubtedly has significant impacts on the labour market, a deeper understanding of the historical and theoretical context of workplace automation proves vital in informing more effective policies for addressing the challenges and opportunities associated with the future of work.

Arntz, Gregory, and Zierahn (2016) directly challenge the methodology and conclusions of Frey and Osborne (2013), proposing a counterclaim in which the share of endangered workplaces is significantly lower for OECD-countries, estimated to be 9 percent. The main point of critique is that Frey and Osborne impute over whole occupations when aiming to assess the risk of automation which overlooks heterogeneity of tasks within occupations, as there are few workplaces which are representative for the respective occupation. Additionally, it is questioned whether the resulting set of 72 occupations at hand of which Frey and Osborne draw conclusions about the overall U.S. labour market is scientifically sound.

As a methodological counterexample they use a workplace-based (individual-based) approach to assess the risk of automation on PIAAC-data for OECD economies, which not only accounts for the variation in tasks within occupations and but also the endogenous potential for workers to adapt to technological change. In doing so, they are able to demonstrate that while potential for automation seems higher when using aggregate task descriptions of occupation taken from O*NET, the same risks aggregated over individuals observed in PIAAC is significantly lower (Arntz, Gregory, and Zierahn 2016). In conclusion, like Autor (2015) they call for less static approaches to assess risk of automation.

2.1.5. Decoding Suitability for Automation

Building on prior research, such as the works of Frey and Osborne (2013), Autor (2015), and Arntz, Gregory, and Zierahn (2016), Brynjolfsson, Mitchell, and Rock (2018) introduce the Suitability for Machine Learning (SML) index, which measures the potential of tasks within occupations to be automated by machine learning technology. The method also serves as an attempt to provide an answer for a problem raised by Brynjolfsson and Mitchell (2017): The fact that no consensus had yet been established on which particular tasks are those in which ML excels compared to humans. For each detailed work activity (DWA) in O*NET, Brynjolfsson and Mitchell (2017) let experts answer a survey containing 23 questions designed to assess a tasks SML. For individuals, the scores then are aggregated per average, weighted by how important a task is for an individual's workplace. Their findings conceptually confirm the claims of Frey and Osborne (2013), that against the backdrop of increasing in AI usage the potential for automation no longer follows the prior pattern of hollowing out the middle, but that automation instead affect all education levels similarly. Yet, in line with Autor (2015) argument about the potential for automation to supplement human labour and to create new opportunities for collaboration and value creation, they discuss the importance of understanding how tasks can be redesigned, emphasising the potential for human-machine complementarity.

Brynjolfsson, Mitchell, and Rock (2018) provided a method to explore which tasks are most suitable for ML but did not yet establish the consensus desired in their prior publication of 2017. Webb (2020) aimed to further contribute to a well-founded, or broadly supported, consensus, by introducing a new method for which he then conducted analysis on the US labour market. In essence, Webb scrapes descriptions and titles of patents in the Google Patents Public Database to identify particular tasks the patented technologies aim to conduct. He divided the technologies into three applications, conceptually resembling the factors discussed prior: Software and Robots, mainly captured by Autor, Levy, and Murnane (2003) and Acemoglu and Autor (2011), and "Artificial Intelligence", which in his case however means ML exclusively; he does not deem programmed intelligence ("sequence of if-then-clauses") to be AI (His reasoning is thoroughly explained in his "Definitions" subsection of the "Application: AI" section). Subsequently, he compared the identified tasks to task descriptions in the O*NET database, allowing to draw conclusions on which are the most prevalent tasks that could - technically – be automated soon and which occupations consequently are exposed most. To assess AI exposure of occupations he, similar to Brynjolfsson, Mitchell, and Rock (2018), uses an average weighted by the respective task intensity.

His findings – to a broad extent – confirm prior theories: Congruent to the job polarisation hypothesis, software indeed mostly affects the middle percentiles of the occupational wage structure, mostly middle-aged men and with a slight decline of exposure with increasing levels of education. The tasks covered by robotics patents mainly overlap with the tasks of male, low-wage and low-education workers, especially with those younger than 30. He predicts decline in wages and employment, of which the former is expected to decline faster. For Artificial Intelligence, he finds a slightly larger overlap for workers with high education, and a steady increase of AI exposure to occupation up to the 85th wage percentile, after which exposure declines. Surprisingly, AI overlaps mostly with middle-aged and higher-aged workers, as well as with men, for which no explanation has been provided before.

2.2. Social Effects of Automation

Providing insights for social policymakers requires the identification of the social consequences automation produces. This section considers two important social effects of automation, namely the hollowing out of the middle, or wage polarisation, and the declining labour share of national income.

2.2.1. The Hollowing out of the Middle

Among many industrialised economies, the disappearance of middle wage occupations is an alarming pattern that can be attributed to a variety of causes, among others technological development and automation. For our purposes, middle wage occupations are understood as the occupations with an income between two thirds and double of the national median income (Kochhar 2015). In the US, the gradual hollowing out of the middle wage segment since the 1970s means that demand for either low- or high-paid jobs has risen in comparison to demand for middle-income jobs, leading to a relative decline in middle-wage occupations. This, for the US labour market, is depicted in Figure 2.1 (Lehner 2017), in which low-wage, middle-wage and high-wage jobs are compared in growth between 2005 and 2016. Hollowing out of the middle initially was an empirical observation, for which economists developed multiple theories. This phenomenon is otherwise also non as job polarisation. Autor, Levy, and Murnane (2003) pioneered to provide automation-associated theory with empirical evidence: The routine-biased technological change hypothesis (RBTC). They identified the diminishing of routine tasks, which often saw their highest share among middle-wage occupations, as a leading cause. For advanced economies, their findings later on were confirmed by many others (Goos and Manning 2007; Goos, Manning, and Salomons 2009; Michaels, Natraj, and Van Reenen 2014)

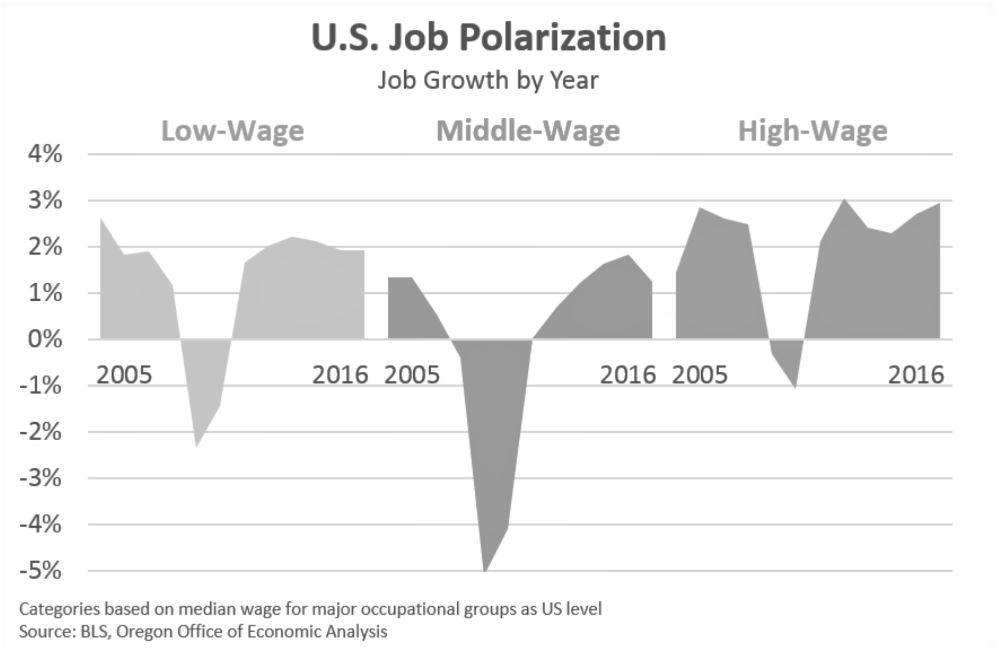


Figure 2.1: Job Polarisation in the United States 2005-2016 (source: Lehner 2017)

Similarly, Marten Goos and Alan Manning examined the growing polarisation of first the British labour market (Goos and Manning 2007) and two years later European labour markets (Goos, Manning, and Salomons 2009). Like Autor, Levy, and Murnane (2003), they find that the composition of the British labour market has changed significantly over the (then) last three decades, with a decline in the demand for middle-skilled employment and an increase in the demand for high- and low-skilled occupations and argue that globalisation, technical advancement, and automation are the main causes of job polarisation in Britain. Later on, the same authors demonstrated the pervasiveness of the RBTC hypothesis and developed a framework to estimate the effects RBTC hypothesis as well as off-shoring. They found that in the general case, RBTC was of higher influence, especially for manufacturing-focused industries, but that off-shoring played a more significant role for service-based industries (Goos, Manning, and Salomons 2014).

In a study associating job polarisation to sinking prices of information and communication technologies (ICT) and research and development investments, Michaels, Natraj, and Van Reenen

(2014) could prove a robust correlation. They offer emprical evidence that the expanding usage of ICT has fueled the polarisation of the labour market by driving up demand for highly trained people and driving down demand for middle-skilled workers. According to Michaels, Natraj, and Van Reenen (2014), ICT has facilitated for businesses to reorganise their production processes, making it simpler to replace ordinary operations with technology while enhancing non-routine duties carried out by highly qualified personnel. By analysing the relationship between the technologies assumed to be the primary cause of the RBTC hypothesis, this further supports the theoretical validity of Autor, Levy, and Murnane (2003).

It must be noted that certain scholars disagree with the notion of RBTC, however. Skott and Guy (2005) critique that concentration of power among occupations, the increasing intensity of effort in lower-paid occupations (Skott and Guy 2007, Crowley et al. 2010) as well as the technology imposed constraints on the freedom of lower-paid workers are not considered within the RBTC hypothesis. An example for instance would be monitoring of workers as seen for truck drivers, which were barely to be located in the late 1980s but usually were satellite-tracked in the late 2000s, or for employees of Amazon warehouses today. They propose the alternative theory of power-biased technological change (PBTC, Skott and Guy 2005): The empirical observation that the effort put into work increased with the advent of computerisation, which RBTC struggles to explain. For the aforementioned examples of truck drivers and Amazon workers, an alternative explanation is provided: That they lost power through monitoring, and that their productivity increased by the newly gained abilities in micromanaging these workers, hence the minimisation of ”unused capacity”. This increase in productivity ultimately renders low-educated workers to be more demanded. Middle-education workers, which suffered from losses in demand moved towards the lower end of wage distribution to account for the shifts in demand. Two years later, a quantitative model has been developed by the same authors, demonstrating significance of their hypothesis by proving PBTC to be capable of explaining demand shifts, real wage decrease and increases in effort intensity (Skott and Guy 2007).

2.2.2. The Great Decoupling: Labour's Falling Share of Income

A prominent observation over the last 50 years is the divergence of key economic indicators across many economies: While GDP and Labour Productivity (Output per hour worked) steadily increased since the second world war, the median household income and private employment however experienced stagnation, for which the term ”the great decoupling” was coined by Brynjolfsson and McAfee (2014). This decoupling is mainly attributed to national income, normally divided in the labour share of income and the capital share of income, incrementally being swallowed by increases capital income which are met with stagnating labour income (Piketty and Zucman 2014). A prominent explanation to this is Capital-Labour Substitution, the effect of how companies and economies decide to rely on capital rather than labour to produce output (Hicks 1932), to be explained more detailed in the following subsection.

Piketty, a luminary in the field of wage inequality and the role of capital in developments of inequality, and Zucman compared developments of the long-term relationship between national wealth and national income through labour. Their findings are represented in his 2014 paper: ”Capital Is Back: Wealth-Income Ratios in Rich Countries 1700-2010”, in which they argue that primary a influence on the fall in labour share of income are the relative gains in capital return. Piketty and Zucman (2014) observed that capital return rates increased most significantly compared to labour return rates in the late 1970s and early 1980s, a time where labour markets were subject to two shocks which represent two forces fueling capital-labour substitution: First and more country-specific, the preferential treatment of capital gains over labour gains by policymakers, for instance through tax systems or weakening of labour regulation. This happened in the early 1980s when heads of states in the largest western economies (Ronald Reagan, Helmut Kohl, and Margaret Thatcher) changed economic policymaking in a capital-preferential paradigm. The second force is technological progress, increases the productivity of capital. A special form of technological progress even more powerful (See Capital-Labour Substitution in Theory and Brynjolfsson and McAfee 2014) is the introduction of new automation technologies, which happened on a large-scale in the early 1980s with computers and later in the 2000s with economy-wide internet rollout (Autor, Levy, and Murnane 2003, Piketty and Zucman 2014).

By the development of a stock and flow model, similar to the field of system dynamics, Hémous and Olsen (2022) managed to explain the role of automation in the decline in labour share qualitatively and quantitatively. Their model in its primary purpose aims to bridge the gap between the priorly discussed macroeconomic observations in association with automation, singular developed theories and the lack of a more dynamic framework which endogenously incorporates creation of new tasks, demand for labour, and progress of automation technologies.

Capital-Labour Substitution

A fundamental concept used by economists to discuss the innerworkings of automation on a macro-economic scale is capital labour substitution. Hicks (1932) first mathematically formulated the substitutional character of the relationship between capital (machines, also incorporating robots, software and AI) and labour in his first book: "The Theory of Wages". Beyond Hicks (1932), many other models exist which all describe these interactions mathematically different, however, rely on similar assumptions. These assumptions are of higher relevance for this thesis than mathematical specifics of these models.

The general mechanism in which capital can substitute labour first requires at least two assumptions to be made, the first being (imperfect) substitutability: To ensure labour can be replaced by capital, capital must be capable of producing the same output as labour can. This, admittedly, is quite flexible in potential interpretations. In simple terms this assumption dictates that the same level of output can be achieved through multiple different combinations of labour input and capital input. And that some labour input can thus be replaced through capital input. (Hicks 1932)

The second assumption – marginal productivity – describes that when increasing only one of the input factors – without proportionately increasing the other – the productivity of the total input productivity of this factor decreases. In other words, capital-intense firms get less output of additional capital input then additional labour input; they are not perfectly substitutable. The factor which describes how substitutable precisely capital and labour are is referred to as the elasticity of substitution. This is relevant, because capital and labour input have costs, which plays a role for the optimisation of profit, which (thirdly) is assumed to be performed by companies under availability of all necessary information. (Hicks 1932)

For capital-labour substitution to occur, it is not necessary to develop automation technologies. Rather, the necessity lies in a productivity-enhancing technology, which mostly would be understood as capital input. This would render capital more productive then labour, which would lead to the profit-optimal input combination to be more capital-intensive, reducing the importance (or necessity) of labour. An automation technology is different from a productivity-enhancing technology in it not only making capital more productive, but also easing the substitution of labour with capital, hence increasing the elasticity of substitution (Tinbergen 1975). This is mainly represented in which tasks a technology can conduct, without having to consider it's cost, which in the case of digital technologies is assumed to be lower than labour cost anyways as digital infrastructure is available. This justifies a fundamental assumption of this thesis: As the combination of ML-enabled automation and routine task automation entails a higher pool of tasks, the change in elasticity of substitution for emerging economies which have not yet experienced computerisation is larger than for advanced economies who experienced distinctly from ML-enabled automation.

Weakening of Labour Market Institutions

Kristal (2013) conducted disaggregated analysis of U.S. industries for the time frame between 1969 and 2007. Her findings revealed that the labour share especially dropped in industries which historically experienced strong union influence that decreased since the 1970s, and which were subject to large scale computerisation. She consequently argues that both factors play a primary role in the power constellation between employers and employees, favouring employers and leading to a decline of the labour share. The two factors, according to Kristal (2013), are not independent but interlinked. She connects both by arguing that the increase in bargaining power of high-wage workers and employers in comparison towards low-wage and middle-wage workers led to a decline in unionisation. For this, she presents two dynamics: First, the fact that computerisation enabled new machines to be constructed to replace routine tasks often carried out by unionised blue-collar workers, which lead to job displacement and reallocation of blue-collar industrial workers from middle-wage employment to low-wage employment against the threat of full unemployment, which not only is in line with RBTC, but also directly weakening union influence. Second, she builds on the uprise of ”technocratic control”, the development that computerisation also facilitated high-detail monitoring of employees and tasks. This in consequence provides employers with a higher ability to exert influence on workers, reducing union bargaining power, and with increased capabilities of granular coordination and reorganisation, increasing their own flexibility and thus bargaining power (Burris 1993, Vallas 1993, Burris 1998, Crowley et al. 2010). This combination of direct and indirect negative influence factors on unions and worker power lead Kristal (2013) to draw similar conclusions about RBTC as Skott and Guy (2005), denoting the relevance of skill and rather depicting technological change as power-biased. Similar to Skott and Guy (2005), Kristal (2013) coined the term as class-biased technical change (CBTC), describing a similar dynamic without shying away from historical connotations.

2.2.3. Policy Responses to Automation-Driven Inequality

In emphasis of labour market changes being not just of technological or occupational nature but also of high social importance, Autor, Mindell, and Reynolds (2022) provides a set of three policy areas with high urgency, in which policymakers can set preparatory measures for upcoming AI-induced labour market disruptions.

The first, taken from the experience gained through prior waves of automation, is education and training to match changing and newly emerging demands. Autor, Mindell, and Reynolds (2022) categorises policies which aim at skills as *supply-side policies*, which creates perspective on how, while education and training policies are in no way useless, they represent only one side of the equation: The management and development of available resources within labour markets. However, he also notes that this is continuously happening and has continuously happened, and by itself promises no match against the (potentially dramatic) changes of future automation waves. In numbers, he used the increase of bachelor-level educational attainment of 75% of young men compared to a 10% real-wage for men over the last forty years.

This rather simple example yet powerfully present the inadequacy of sole reliance on the supply-side when aiming to maintain a ”fair” labour share of income, which accordingly justifies the necessity for demand-side addressing policy area two: Investment into institutions of various forms (state-owned, free-market based, inter-govermental) which put effort into distributing income fairly, as a fair response to productivity increases (Autor, Mindell, and Reynolds (2022) chose the formulation: ”translation of productivity gains into shared prosperity, irrespective of failure or success"). It is noteworthy to mention that the main factors to which the divide in income share is attributed to mainly are policies which favour capital gains over labour income and no market forces (Autor, Mindell, and Reynolds 2022). Like Acemoglu, He, and Le Maire (2022) and Autor (2022) (and many others), he credits shareholder capitalism with its contribution to productivity gains and wealth creation, but especially in light of automation warns of its limitations in assuring that creation of value also leads to improvements in life. In more concrete terms, he lists (often US-specific) policy alternatives such as the update and stricter enforcement of labour standards, minimum-wage policies, more extensive unemployment insurances, and changes to the (often) employer-based health insurance system (Autor, Mindell, and Reynolds 2022).

The third policy area is the intervention into how technology develops in productivity and skill complementarity; efforts to design technology beneficial to society rather than individual players. In particular, the development or strengthening of public innovation and R&D system to keep public interests reflected in technological advancements. Autor, Mindell, and Reynolds (2022) argues a more participatory approach fosters the potential to ensure diffusion of AI on a more desirable path. When AI is developed to aid workers instead of replacing them, the balance of productivity gains and maintenance of employment and wages rather is in favour of workers. Moreover, participatory approaches prove great efficacy in preventing harmful consequences of AI-systems as those affected are also brought to the table (D'ignazio and Klein 2020, Dobbe, Gilbert, and Mintz 2021).

2.3. The Case for Emerging Economies

2.3.1. Routine Task Automation in Emerging Economies

Routine tasks, automatable through computers or robots (Autor, Levy, and Murnane 2003, See Chapter 2.1.2), currently are still the major focal point in the scarcely available academic literature on automation in emerging economies (Messina 2016, Maloney and Molina 2016, Das and Hilgenstock 2018, Carbonero, Ernst, and Weber 2020 Lewandowski et al. 2022). Research on automation through AI is especially scarce (Carbonero, Davies, et al. 2023).

Another relevant and more frequently discussed point however is that with increasing globalisation, many multinational companies offshore their production centers from developed to emerging economies. This shift is motivated by multiple factors: competitive advantage in labour costs, new consumer markets, gaining beneficial conditions through subsidies (Porter et al. 1998). Among many reasons, competitive labour cost advantages compared to developed economies but also to machines in emerging economies, let Autor, Dorn, and Hanson (2013) argue that routine task intensive occupations actually increase in emerging economies as they get offshored by multinational companies. This could be empirically shown for China (Du and Park 2018) and some eastern European countries (Hardy, Keister, and Lewandowski 2018), however is not yet shown a global scale.

Studies which analysed routine task intensity in emerging economies did not reach alignment over the decline of routine intense occupations of Autor, Levy, and Murnane 2003 so far, showing neither that occupations with high routine task intensity declined in an economy-wide magnitude (which partially supports the hypothesis of Autor, Katz, and Krueger (1998)), nor that they characteristically match those with mid-level education or wages (Messina 2016, Maloney and Molina 2016, Das and Hilgenstock 2018). In fact, the circumstance that educational attainment in emerging economies typically is considerably lower in emerging economies than in developed economies justifies the expectation that many occupations are matched with workers with lower educational attainment. Under the assumption more low-educated workers exist, routine jobs would rather be allocated to workers with low education (Messina 2016, Das and Hilgenstock

2018).

However, despite exposure to routine tasks having increased over time in some emerging economies without measurable changes in occupational structure, Das and Hilgenstock (2018) do not reject the automation of routine tasks in emerging economies, but instead argue that automation has not taken place yet (as of 2018), or has been impeded in measurement through influx of employment through offshored routine employment. However, Das and Hilgenstock (2018), like Lewandowski et al. (2022) in more recent research, could demonstrate that with increasing technological progress and availability, routine task intensity was lower. Moreover, both Das and Hilgenstock (2018) and Lewandowski et al. (2022) found technological progress to be of higher weight than the level of globalisation, implying a high future potential for routine task automation.

2.3.2. AI-enabled Automation in Emerging Economies

Beyond Carbonero, Davies, et al. (2023), the literature on AI-enabled automation in emerging economies is scarce, and no further sources are known, rendering the state of affairs concerning AI-enabled automation in emerging economies ambiguous. Insights on AI automation potentials will be provided by this thesis' quantitative part. However, their educated interpretation – beyond which occupations have high values and which occupations don't – requires an understanding of the institutional context around their adaptation. For emerging economies, the past has brought up numerous examples of market dynamics in which new technologies introduced by companies from advanced economies with high know-how. The results were often times nuanced, resulting in productivity gains and economic growth on the one hand, but fell short of ensuring the gained wealth benefitted the citizens of emerging economies to which new technologies diffused.

This section provides the tools for an educated interpretation of AI-automation indicators in emerging economies, to be used in the later discussion of policies (See: Chapter 5). First a model technical diffusion is provided, conceptualising the life-cycle of the most important technologies, within economies. Expanding the scope to multiple economies with different stages of development, concepts of developments economics shown among historical examples are brought in, at hand of which potential dynamics behind the diffusion of AI into emerging economies are discussed.

Life-cycle of General Purpose Technologies

General Purpose Technologies (GPTs) are defined as technologies which possess three properties: their usage is spread across a variety of industries, they foster innovation in their application, and they facilitate efficiency increases. Only few technologies count as GPTs, as they are attributed to bring unforeseen societal and economic change. Examples of GPTs are the steam engine, electricity, and the internet (Bresnahan and Trajtenberg 1995).

Perez (2003) developed one of the most influential models of diffusion and impact of GPTs. According to Perez, the life-cycle of GPTs can be divided into two main phases: the installation and the deployment phase, as depicted in Figure 2.2. Each phase is further divided into two sub-phases.

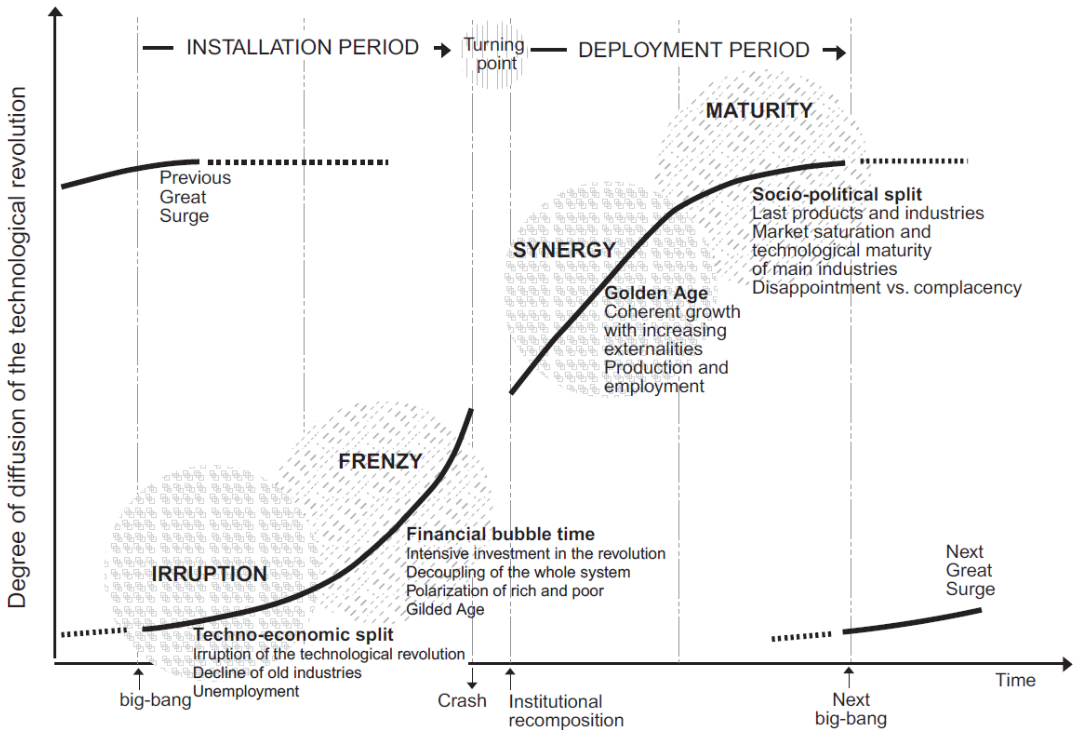


Figure 2.2: Life-Cycle of GPTs (source: Perez 2003)

Installation Phase

Irruption: The irruption phase marks the advent of a new technology. This is when a groundbreaking innovation happens and the first models or iterations of the technology are created. Typically, technologies are still in their infancy in this stage, and most prospective applications are yet to be identified. Technologies are frequently met with scepticism and confusion. Because of this, only a small share of potential investors are ready to commit to further development (Perez 2003).

Frenzy: As financial capital begins to acknowledge the potential of new technologies, investments soar in the frenzy phase. Investments in businesses and startups associated to the technology are in high demand. Speculation, excessive investment, and potential bubbles are characteristic to this stage. During this phase, competitiveness and first-mover advantages are prioritised, and social considerations frequently disregarded. (Perez 2003).

Deployment Phase

Synergy: The emphasis switches from competition to production during the synergy phase. The technology has advanced considerably now, and many field of application have been identified. The advantages of the technology are understood widely, initiating the spread throughout many industries and sectors. (Perez 2003).

Maturity: Technology reaches saturation during the maturity phase. It is now frequently employed and integrated into the social and economic structure. The rate of technology returns starts to slow down, and incremental innovation rather than radical change becomes more important. (Perez 2003)

Latecomer Advantages, Catch-up Effects, and Leapfrogging

In economies in which GPTs originate, typically advanced economies, the installation period per definition takes as long as it requires (prevailing) players to gain understanding of a technologies' potential uses and its integration into business processes (Perez 2003). Typically, such development requires iterative improvements in the technology itself, its use and consistent discovery of new fields of application. The further these iterative processes let a GPT mature in an advanced economy, the more steps can be skipped by economies who have not yet adopted the same GPT. This is what economic historian Gerschenkron (1962) postulated as the latecomer (or backwards) advantage. Inspired by the catch-up of the Russian Empire to industrialised Continental Europe, Gerschenkron (1962) argued that the latecomer advantage provides a unique opportunity for economic catch-up through imitation of technological progress. Latecomers can skip over different stages of technological development and accelerate their economic growth by adopting General Purpose Technologies (GPTs) that have developed in advanced economies. Yet, Gerschenkron (1962) posits that every economy that adopted late will invariably add a note of its own national characteristics to the adoption process, leading to different developmental paths. These nuances can result from a broad range of factors, including local market demands, governance structures, or cultural attitudes towards technology. Despite these variations, the degree of backwardness nonetheless remains the primary driver for economic catch-up (Gerschenkron 1962).

Beyond the benefits to national economic growth, the path of adoption deserves critical attention. In many cases in the digital economy, large players hold licensing or intellectual property rights, which for a limited timeframe provides them the right to control the distribution and use of a technology (Perez 2003). This power becomes problematic for other economies in which a GPTs installation phase begins later than the begin of the deployment phase of the economy in which a technology originated. Since advanced economies have contributed significantly to the knowledge and development of a GPT, the GPT may diffuse into emerging economies at a significantly riper stage, with players who already have a very advanced know-how, potentially even holding monopolies. Some emerging economies may bypass entire stages of technological progress as a result, rendering it close to impossible for national firms to keep up (Iyer 2018; Homsombath 2020).

The consequences of labor-saving effects may be exacerbated further by this leapfrogging through two channels. First, rapid import and adoption of AI technologies may not give the local workers the time to adjust or develop the required skills. In the installation period in developed economies, the technology itself has not only had the time to reach maturity, but business models emerged out of the necessity to introduce and harness a technology most effective. Second, it is likely that the players who own the software or intellectual property rights and who license out the their technology will end up with a disproportionate share which they take back into advanced economies, partially depraving emerging economies of technological benefits. One example of an innovation which imposed such twofold harm on emerging economies is the mechanisation of coffee farms: The demand for human labour was significantly reduced as automated coffee harvesting machines were introduced, but at the same time, the gains made from enhanced efficiency were mostly returned home by the foreign firms that controlled the technology. Due to this, there were fewer job prospects and little to no local economic benefit from the higher revenues. (Juma 2016).

Gerschenkron (1962) and Perez (2003) underline the need for emerging economies to take the initiative in creating plans to take advantage of a new technologies' benefits while minimising its negative impacts. This involves making educational investments, supporting regional innovation, and creating legal frameworks that enable inclusive growth. Anticipating the economic and societal effects of GPTs, such as AI, requires an understanding of their life-cycle. At it's current stage, nearing deployment two developments must be anticipated: Its labour-saving nature can lead to wage suppression and job losses in the short term, but it also has the potential for job creation in the long term.

2.3.3. Policy Considerations for Emerging Economies

Routine Automation Implications

Just as for developed economies, the implications of routine task automation are nuanced. Desirable effects are economic growth through efficiency increases, reduction of errors, and benefits in scalability. This could decrease the cost of goods and services for consumers and possibly boost the emerging economies' global competitiveness. Also, gaps within the labour force could be filled, relieving emerging economies of disruptions through dependencies on scarcely available skills.

On the other hand, the widespread automation of routine tasks can pose challenges for the labour force. In emerging economies, routine employment predominates in many industries, including manufacturing, data entry, and basic customer service. Machines replacing these jobs may cause short-term increases in unemployment and force the labour market to change its requirements for skills. It may be necessary for workers to retrain or upskill in order to stay relevant, and educational and vocational systems urgently need to adapt to this shifting environment (Autor, Mindell, and Reynolds 2022). Additionally, automation's advantages are likely to be distributed unevenly. Workers who are replaced by automation may experience income losses and financial troubles, whilst companies and customers may benefit from cheaper goods and services (Autor, Mindell, and Reynolds 2022). As individuals who own and manage the tools of automation reap a bigger proportion of the economic advantages, this may worsen income disparities within emerging economies (Juma 2016, Das and Hilgenstock 2018).

In conclusion, routine automation creates chances for economic development and efficiency in developing nations, but likewise fosters problems that require urgent action upon appearance. To minimise harm during increasing automation, inclusive and advantageous for all facets of society, policymakers, educators, and company players must operate in a way that enables economy-wide welfare gains, which is rather unlikely to be facilitated by unguided technological diffusion.

Differences and Augmentations to Policies for Developed Economies

Given the lack of available literature, this subsection will synthesise and discuss insights of prior subsections to provide a picture on what caveats need to be considered for policy-making aimed to prepare labour markets in emerging economies for automation. Further inspiration for this chapter is taken from two sources: First, the discussion of policies preparing for automation in advanced economies by Autor (2022), whichs' proposals are evaluated against characteristical differences between the labour markets of advanced and emerging economies themselves, and the diffusion of technology in advanced and emerging economies. For further augmentations, inspiration is taken from Dutz, Almeida, and G. (2018) who compiled a summary of insights for labour market transformation in Latin and Central America, also in the light of technical change, for which they provided policy recommendations.

While the policy areas as discussed by Autor, Mindell, and Reynolds (2022) generally provide and understanding and sense of direction for what policymakers need to account for when designing sets of measures preparing for automation, their ideas mainly focus on advanced economies. For emerging economies, important distinctions must be made, of which several were discussed in the preceding subsection:

The first two effects affect how automation technologies enter labour markets, which impedes potentials to be used in the R&D-policy area Autor, Mindell, and Reynolds (2022) discussed. Starting with the effect of automation technologies arriving at a riper stage with a established ownership structure concerns emerge who technologies aim to benefit. It is not to be expected that emerging economies which have lower output in automating technologies, lower know-how and little resources to allocate towards R&D will prevail in competition with the innovation systems of countries like the US. A second threat to the usefulness of R&D policies in emerging economies is the potential of multiple automation technologies (AI and routine automating technologies) to reach large-scale application at overlapping or even the same times; when digital infrastructure does not just open the door to information and communication technologies but also immediately to sophisticated AI-models, labour market effects are likely to be more powerful than with singular arrivals of technologies. These overlapping arrivals of technologies can make the trajectories of technical diffusion more complex and can lessen the influence national innovation systems will have on new market dynamics altered by availability of new technology. Overall, policy potential of interventions into how AI will reshape emerging economies is expected to be far lesser than for advanced economies. New routes must be explored, like the leveraging of diplomatic ties or state-programmes steering the diffusion of technologies, entailing government investment in directed rollouts of digital infrastructure sidelined by digital literacy programs, potentially at the expense of higher government expenditure and not realising the full potential of on-paper productivity increases.

Third, labour market informality hinders efficacy of policies on labour markets as compliance to and enforcement of the policies is inhibited by the lack of knowledge on informal sectors. This might impose difficulties the ”demand-side” area of policies as discussed by Autor, Mindell, and Reynolds (2022): In economies with large informal sectors, only little reliable data can be obtained to understand how new technologies would cause decreases in the labour share. Effects of inequality through automation are hence not just difficult to anticipate, but can also hardly be assessed in retrospect. Policies attempting to steer the distribution of resources more fairly during and after automation thus will inevitably have design flaws, and on top can hardly be refined iteratively. In the likely case of increasingly automated labour markets producing increasingly less welfare for workers, systematic policy approaches as opposed to symptomatic policy approaches would be subject to such high uncertainty, seriously challenging systematic policies aimed to be robust. However, symptomatic policy approaches (such as government handouts to those hit by unemployment and subsequent poverty) will lead to budget difficulties in the long run if no beneficial systemic changes occur.

Fourth, the ”supply-side” policy area as discussed by Autor, Mindell, and Reynolds (2022), the classical proposition of educational or vocational adjustment, is challenged due to comparatively weak education systems. Not just is the labour force educated lower on average, but also the infrastructure for educational adjustment is less present and has to operate on lower resources. Potential for skill attainment increase with higher availability of digital infrastructure through means provided by private companies like Coursera, but can at best only level the playing field in educational availability, and cannot fully compensate for shortcomings of the general educational system.

Dutz, Almeida, and G. (2018) propose – next to supply-side policies – two further policy areas to steer economic developments in the face of technological advancement for Latin and Central America. Two aspects must be noted: First, these policies do not exclusively focus on automation technologies, as this thesis does, and secondly the perception of what is desirable differs slightly. While this thesis welcomes economic growth but aims to assess who this growth benefits, Dutz, Almeida, and G. (2018) primarily focus on inciting economic growth, understanding growth as a prerequisite of inclusive growth. The first policy area proposed by Dutz, Almeida, and G. (2018) tackles technical diffusion. In particular, aiming to leverage productivity potentials through digital technologies, internet rollout is mentioned as the core lever to facilitate technical diffusion. Emphasis is set on shaping markets in a competitive fashion to drive the prices of digital infrastructures down and increase firm-level adoption. Another lever mentioned is the reduction of tariffs on ICT services, which for the region of discussion are extraordinarily high, attributed to delays in ICT-adoption (Dutz, Almeida, and G. 2018). The implications of slowed adoption are ambivalent. While output growth is undeniably inhibited, policymakers and other context setters also have more time to design the grounds of interaction in a way preventing exploitational market dynamics, as described by Juma (2016).

Second, Dutz, Almeida, and G. (2018) bring in how product market policies, policies that are aimed to improve the environment of competition within markets, can help inclusive job creation. In particular, they suggest that if businesses have to compete more fiercely, they are more likely to innovate and grow, which can result in more job opportunities. This is especially true in markets where businesses can easily enter or exit, and where there's less risk involved in expanding operations. One proposed product market policy for sectors where competition and technical adoption is low is the easing of financing for technical adoption for less-established firms can prevent "winner takes it all"-dynamics and instead foster growth of small local companies. Beyond financing for technical adoption, Dutz, Almeida, and G. (2018) propose policies strengthening the know-how of management in local firms. This is aligned with Gerschenkron (1962), who stressed that with higher backwardness the strength of institutional guidance must increase to ensure a seamless adoption. The overarching objective of product market policies would be to invigorate the business landscape, thereby tailoring market conditions to promote stable and inclusive employment opportunities.

In summary, the potential of policy intervention in emerging economies is challenged compared to advanced economies, which most likely will complement challenges induced by lower resources, more complex dynamics of diffusion and labour regulation (which was purposefully not discussed prior given the diversity of labour protection policies in emerging economies) in labour markets. Contrary to advanced economies, policymakers might be forced to resort to a broader policy set, exceeding the space of labour market policies, and might be struck by larger arising consequential complexity, and more diverse consequential problems.

Methodology

An explanatory sequential mixed-methods design, as described by Creswell (2020), will serve as the foundation for the research for this thesis. The identified indicators will be calculated at first on individual level after which they will be aggregated to occupations for the STEP data of emerging economies. For RTI (Autor and Dorn 2013), calculations are conducted at hand of the direct data sources. The calculation of SML is based on task descriptions in O*NET (Brynjolfsson and Mitchell 2017). As we compare data of countries for which no O*NET data is available, we will apply a methodology introduced by Carbonero, Davies, et al. (2023) utilising the SBERT-language model. For this, we will first create similarity matrices between task descriptions in O*NET and STEP, and then calculate SMLs based on task prevalence averages, weighted by the similarity to and the SML-score of detailed work activities (DWA) and general work activities (GWA) in O*NET.

Hereafter, occupations will be selected according to data abundance in both data sources and overall comparability between emerging and developed economies. The selected occupations will subsequently be compared per indicator and analysed for reproducibility of the insights obtained in the literature review. The most vulnerable and at-risk groups within the labour markets of developing economies will be identified after the data has been gathered using cluster analysis. Cluster analysis is chosen as it constitutes a machine-learning based classification task, which does not require supervision. Supervised classification techniques would require a precise labelling of whether risk of automation is given or not. This, to some extent, could be conducted at hand of the calculated indicators. However, it is currently unclear according to which framework these indicators can be used for labelling, and development of a framework is deemed as not feasible within the scope of this thesis. It was alternatively decided to use the indicators for post-hoc analysis. Clustering will be conducted exclusively at hand of socio-demographic factors. A task-based clustering is deemed as inviable, as the indicators used in post-hoc comparison are calculated at hand of task-based variables and could thus infer explainability issues.

3.1. Data

The indicators on risk of AI-enabled automation by Autor et al. Autor, Levy, and Murnane (2003) as well as Brynjolfsson et al. Brynjolfsson and Mitchell (2017) are based on what tasks are conducted within particular occupations. This section will describe and motivate what sources of data were selected, which data is contained on task and skills within occupations, and what socio-demographic data can be taken from our sources for further analysis.

3.1.1. Data Selection

The most influential methods, namely of Autor and Dorn (2013), Frey and Osborne (2013) and Brynjolfsson and Mitchell (2017), used task data taken from O*NET, a database which encompasses the worlds most extensive range of factors that reflect job and worker characteristics, such as skill requirements and activities performed, but only covers the US labour market. Given the data quality prior research has been conducted on, a need for data on other economies, irrespective of development status, emerges with high demands to data abundance (number of observations) and comprehensiveness (depth of observations). Simultaneously, concessions must be made to the fact that alternative data sources cannot be expected to hold up to the standards of O*NET. Additionally, to ensure procedural feasibility and validity of analysis emphasis is put on data being consistent or, when taken from different sources, comparable for all countries analysed.

Within the search for data sources, no single database which contained data on a diverse set of economies could be found. However, data containing detailed task information structured alike and thus comparable over multiple emerging economies could be found: The STEP-survey issued by the World Bank (World Bank 2014). Accordingly, STEP will serve as the primary sources of data for this thesis. In the following two subsections, STEP will be described, after which further identified data sources which will not be considered are introduced along with the reasoning behind their exclusion.

3.1.2. STEP

The World Bank's STEP Skills Measurement Program (STEP) is the "first ever initiative to measure skills in low and middle-income countries" World Bank 2014. The two STEP surveys, the household and employer survey, collect data on their respective subjects. The household survey contains information on three elements: demographics of households, competences, psychological traits, and work-relevant skills. The employer survey includes five: labour force structure, cognitive abilities, personality traits, and job-relevant skills that are currently in use; skills employers look for when hiring new employees; the way employers provide training and compensate their employees; and the degree of satisfaction with the level of education and skill training available in the labour force. (World Bank 2014)

To calculate SML and RTI, which both mostly require information concerning tasks and skills on first individuals and then occupations, the household survey proves most useful in providing us the required data. The household survey is comprised of nine modules: The first module contains socio-demographic and residential information on a household level, which is defined as as a group of people who live together and make common provision for food or other essentials for living (World Bank 2014). This module also plays an important procedural role, as within part C, one individual of the household is chosen to participate in the subsequent surveying process, for which a new appointment was made with the survey participants. In consequence, the STEP survey contains only one observation per household, of one randomly chosen individual. For these observations, modules 2 to 7 contain data on individual education (module 2), health (module 3), employment (module 4), work skills (module 5), psychological assessments (module 6) and family situation (module 7). Module 9 contains a standardised assessment of reading capability. Modules 8 and 10 are filled by the interviewer and contain their own perceptions of the survey process with the respective individuals. (World Bank 2014)

For this thesis' purpose, three modules are of particular importance: Module 1 provides the data necessary to conduct a socio-demographic cluster analysis. The precise choice of variables with justification is to be found in Section 3.4. Furthermore, module 1 contains data on how representative individual observations are in the form of weights, which are used for this thesis' visualisations of data, standardisation of automation indicators, conduction of socio-demographic cluster analysis and training of regression models. Module 4 provides data on employment status, wage or pay, and as such also plays a crucial role in the cluster analysis. Module 5 facilitates the calculation of automation indicators. It contains 134 survey items on tasks or skills, which in the STEP survey are understood synonymously. For instance, a general skill or task such as computer use is inquired as whether it is required on the respondents job, irrespective of whether the respondent is capable to use computers. If a respondent answers positive on the inquiry of general skills, the follow-up questions inquire how computers are used in detail, i.e.

for writing e-mails or for programming, and how often the skill is required. This provides a task description which more closely resembles the occupational skills of an individual rather than the individuals skill set. However, not all survey questions within module 5 are directly associated to specific tasks (Do you sometimes spend more than 30 minutes to think about problems?), or often only to specific tasks with no overarching general tasks (Do you use a bar code reader?). Consequently, not all questions are suited for a standardised assessment of tasks on the job, constituting the need for a subset of survey items with direct task-association. How the subset is created differs between the indicators we calculate, and is explained in detail in the respective indicators sections (See Section 3.2 for RTI and Section 3.3 for SML).

3.2. Routine Task Intensity

The assessment of Routine Task Intensity (RTI) traditionally follows the definition of routine tasks being tasks which underlie a set of well-defined rules, which software or a simple robotics could perform, as in Autor, Levy, and Murnane (2003). Non-routine tasks are tasks which are only to be conducted by humans or more advanced technologies. While the term has been coined as ”Routine Tasks”, the question whether a task is routine, as in repetitive or something a worker is used to, actually is only of minor importance. What the measure rather aims to describe is whether a task can be automated by computers or robots, as a task can be conducted again and again by always performing the same steps under a given set of circumstances (which can be expanded to a large set of circumstance sets).

A further distinction lies between cognitive, manual and interactive tasks. This enables us to not only distinguish between activities that computers or robots could or could not carry out, but also to take into account the kind of tasks carried out, providing us with a richer picture of the work we examine. The categories resulting from the combination of non-routine or routine and manual, cognitive or interpersonal tasks (such as non-routine cognitive or routine manual tasks) are called composite task measures, following the definition of Autor, Katz, and Kearney (2008). Composite task measures themselves comprise individual tasks. Applied to STEP, this means that individual survey items containing task descriptions were analysed and subsequently allocated to either one or multiple composite task measures. Ultimately, the RTI-index represents the result of weighting routine composite task measures against non-routine composite task measures.

It is furthermore important to note that RTI, just as its composite task measures, is measured on an ordinal scale. Autor, Levy, and Murnane (2003) devised the calculation in a way which facilitates comparison between groups of low and high RTI, but have not envisioned a particular meaning behind the actual numeric values or differences between them. Rather, RTI, serving the purpose to explain automatability, is appropriately distributed if for a given set of occupations, those who reflect the theoretical definition most are on the other extreme of the distribution as those occupations which disalign most with the original definition.

3.2.1. Calculation Procedure

The calculation of an RTI-index consequently requires the following steps: First, composite task measures must be chosen, to which there are two available alternatives: The extensive categorisation follows the definition of Autor, Levy, and Murnane (2003), hence dividing into routine cognitive, routine manual, non-routine interactive, non-routine cognitive and non-routine manual and has been used by Spitz-Oener (2006), Black and Spitz-Oener (2010), and Acemoglu and Autor (2011). Faced with data allowing for less sense of detail, Autor and Dorn (2013) devised the aggregate RTI-index, comprised of routine, abstract and manual tasks, of which the latter two contributed negatively to RTI. The aggregate version of RTI was later on used by Autor and Handel (2013) and Goos, Manning, and Salomons (2014). Others, like Firpo, Fortin, and Lemieux (2011) or Green (2012) reject the categorisation into either routine or non-routine and use deviating task description indices with different composite task measures, which are computed analogously, however. Within this thesis it was decided to follow the mainstream of literature, using the aggregate and the extensive RTI, to evaluate against the available data in later steps.

Second, individual data points, which in sum constitute the composite task measures, must be selected and categorised in accordance to the chosen composite task measures. Because the individual data points are usually summed in order to calculate the composite task measures but have different value intervals, the individual data points are normalised, initially providing each data point the same weight. This step required extensive conceptual work on the individual survey items of STEP, which is described in the subsections *Obstacles with Definitions and Data* and *Overcoming Data and Definition Obstacles*. The results are displayed in Appendix B.

Third, a decision on whether and how individual are task measures are weighted has to be performed. Options on the "how" include principal component analysis (PCA) or manual evaluation of perceived importance. Both options have benefits and drawbacks; so does PCA reduce overlap between variables and would lead to very high emphasis of task items which have low correlation to others, but would also emphasise these task items if they have lesser descriptive potential, which in turn could skew the weighting. On the other hand, can manual analysis allow for higher inclusion of theoretical concepts but simultaneously be prone to bias of the author. For composite task measures, standardisation is chosen over normalisation as per the following reasoning: While we want to prevent individual measures from dominating and ensure comparability over all task measures of the same kind, composite task measures already are aimed to be descriptive, just as the final RTI-index. A focus on a central tendency (i.e. a mean set to zero) and a standard deviation of one facilitate clear interpretation of composite task measure values. This is turn further increases the meaning of our RTI, which is aimed to be a relative measure of "routineness" interpretable towards other occupations, not towards the maxima.

Fourth and finally, the choice of formula for an RTI-index remains. Within the literature, two alternatives can be found. Either are composite task measures summed up straightforwardly (i.e. Autor and Handel 2013), or are logarithmically transformed before (i.e. Autor, Katz, and Kearney 2008). Logarithmic transformations in economics are traditionally applied for one of two reasons: Either to mitigate skew or outliers, which our data has been freed from in advance per normalisation and standardisation, or to make-up for multiplicative or exponential effects, which do not occur within our composite task measures. Accordingly, the RTI calculated in this thesis follows the simple sum equations as depicted in Equation 3.1 and Equation 3.2.

$$RTI_{aggregate} = CTM_r - CTM_a \tag{3.1}$$

$$RTI_{Extensive} = CTM_{rc} + CTM_{rm} - CTM_{nrc} - CTM_{nrm} - CTM_{nri} \tag{3.2}$$

Where r stands for Routine, a for Abstract, rc for Routine Cognitive, rm for Routine Manual, nrc for Non-routine Cognitive, nrm for Non-routine Manual, and nri for Non-routine Interactive.

3.2.2. Obstacles with Definitions and Data

Certain friction points during the application of the RTI-method from O*NET onto STEP deserve attention, as they are either endangering the quality, depth or suitability of analysis. First, the original RTI-calculation by Autor and Dorn (2009) was performed on the O*NET-database, out of which 16 task descriptions were picked, which in sum constitute the representation of 5 composite task measures, for each occupation. For STEP survey items however, the imposed standard of clarity could not be met, as instead of a highly granular set of occupation- and context-specific questions, 47 general questions on work activities were asked on an individual rather than occupational level. The vague character of questions naturally makes it more difficult to categorise into routine in non-routine, which in the later stages of our analysis will impede our capability to measure routine content. This issue is in the following referred to as the ”measurability issue”.

The measurability issue also is further exacerbated by the lack of importance weighting and the lack of frequency indications for many survey items in STEP. Consider the survey item ”Do you use a bar code scanner at your work”. Without a frequency important indication, the same data point would be generated for workers for which this item plays a clearly different role. For three exemplary workers, one supermarket cashier, one Amazon warehouse worker, and one supermarket manager, the same datapoint would describe three different realities: For the supermarket cashier, who uses the bar code scanner frequently and finds his core task represented in this question, the automation of a cash register – as seen in many advanced economies – means the automation of their occupation. The warehouse worker would in a more modern warehouse, like most amazon warehouses, scan the bar code of an object to provide data about the inventory, the status of his own task, and the status of an order to the Amazon system. If, say through an intelligent inventory system, the scanning of bar codes gets automated, they would still bring objects from A to B, prepare orders, and conduct other tasks. The task of scanning bar codes occurs frequently but is not as crucial, it’s automation would be convenient to the worker and potentially slightly productivity increasing. For the supermarket manager, scanning bar codes is neither crucial, nor does it occur as frequently, yet would they likely respond with ”yes”. For the calculation of RTI (and SML likewise) the answer to the questions would be incorporated equally for all three individuals.

Moreover, conceptual exceptions emerged which demonstrate that the definition provided by Autor, Levy, and Murnane (2003) fails to account for more instances of digitalisation ”evading” certain tasks but still reaching the goal of these tasks. Tasks, for which certain steps require human capabilities, could be not automated but replaced by computers as the tasks were abstracted from start to finish: One suitable example is the task of reading forms, or other documents comprising structured information. While the recognition of individual hand-written entries in a form, or a signature, initially was deemed as non-routine, in-practise digitalisation proved reading forms can nonetheless be replaced through computers. That is as in many workflows forms serve the purpose to inquire and save data, to be used for later data entry and then processing. Digitalised forms, such as structured PDF-documents or services like Google Forms, can serve as input documents and facilitate the immediate transfer of their data to the processing software, automating the work of numerous civil servants. These civil servants could be replaced because the tasks they conducted were made unnecessary through technological process even though their solutions, in our case digitalised structured forms, cannot perform the direct tasks required, such as reading a paper document. In the case of reading forms, an inherently instrumental task for follow-up tasks such as data entry, an immediate bridge from data collection to data processing could be build, cutting the need for data entry and hence for reading forms. For our analysis, this means we must conceptually differ from the original definition of routine tasks provided by Autor, Levy, and Murnane (2003), if we aim to account for the previously described replacement effects. This issue is in the following referred to as the ”conceptual issue”.

3.2.3. Overcoming Data and Definition Obstacles

Autor and Dorn (2013), acknowledged the issues in measurability of routine tasks when attempting to bring his concept from Autor, Levy, and Murnane (2003) to the data. Out of all potential origins of measurability problems he identified, the case of STEP, luckily, is one of the least troublesome. For instance, it must be noted that the design of surveys which particularly aim to assess routine task intensity often encounter in-practise issues, as survey respondents often have intransparent and varying understandings of what would be an automatable routine task and what would not be. He uses the example of a cleaner as a survey respondent, who is unaware of his perceptional advantages towards a potential cleaning robot, and thus would falsify data provided on routine content. In our case, where the data used does not include the specific context of tasks, we gain the advantage that: ”*Adding task measures directly to worker surveys places no restrictions on the variability of tasks within as well as across occupations, does not inadvertently impose the assumption that occupational tasks are static, and allows task measures to be designed for testing specific hypotheses*” (Autor and Dorn 2013).

Despite the survey items of STEP requiring manual categorisation, the measurability issue therefore does imposes no severe limitation; as long as the manual categorisation can accomodate for vagueness and pitfalls can be successfully avoided.

Two pitfalls could be identified: The first is miscategorisation or non-categorisation induced by lack of acknowledgement of the conceptual issue. Given that tasks are described unspecific to context and are scarce in numbers, too strict categorisation after Autor, Levy, and Murnane (2003) could lead to valuable information being ommited, e.g. if our example of ”reading forms” were to be ignored or categorised as ”non-routine” because reading forms, in a strict sense, requires human vision too take in information. The second pitfall is the overconfident conclusion that because a task could not be automated but replaced through computerisation, it immediately qualifies as a routine task. The circumstance that the vague task descriptions in STEP can be representatives of many similar, but yet different tasks, forces us to ensure that a task at least in the majority of cases is a routine tasks before categorising it as such. When not overlining clearly with the conceptual vision of Autor, Levy, and Murnane (2003), a task can be deemed as a routine task only if a task description seems generally automatable or replacable through computerisation.

Admittedly, this cannot be conducted to perfection and is vulnerable to omission of cases in which a task can both be replaced and not be replaced dependent on the specific circumstances. Yet, a categorisation process wary of the pitfalls identified priorly, if incorporated schematically, allows for a more nuanced assessment of routine task intensity, reducing the likelihood of incorporating author bias. To maintain transparency and logical consistency, a set of four criteria, one of Autor, Levy, and Murnane (2003) and three aiming to compensate for our pitfalls, was conceived before categorisation into routine and non-routine.

- Criterion 1 (Rule-based character): Can the described task directly be translated into a set of steps which are performed dependent on pre-defined conditions?
- Criterion 2 (Abstractability): If no, can the described task be abstracted into initial state, goal state and a different sequence of steps, to then be carried out by a machine?
- Criterion 3 (Generalisability): If criterion two is met, is a hypothetical substitute of this tasks generally applicable or ambiguous?
- Criterion 4 (Informative Value): If ambiguous, does the task provide other information indicating substitutability by software or robots?

These four criteria initially follow the logic of Autor, Levy, and Murnane (2003) in criterion 1, dividing tasks into routine and non-routine tasks, by asking ”can the task be translated into a

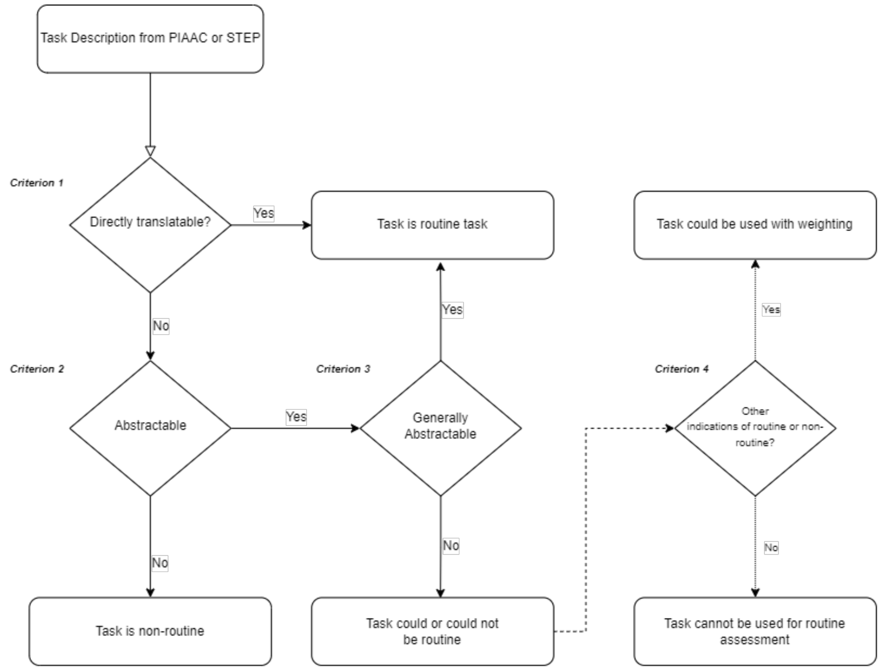


Figure 3.1: Categorisation Criteria as a Sequence of Yes-or-No Questions

well-defined sequence of steps with rules?". Hereafter, our pitfalls are considered. Criterion 2 was framed so that questions as "Do you (as a regular part of work) read forms?" could be accounted for. Yet, not every task description which could be abstracted to a process with a start and a finish equates a task that also can (with certainty) be performed by robots or computers. Tasks for which it seemed ambiguous whether abstraction could be generally applied to were identified with Criterion 3 and with the help of criterion 4 were deemed as inconclusive, and hence not considered further. The criteria, as a sequence of Yes-or-No questions, are depicted above in Figure 3.1

After both the aggregate and extensive RTI had been chosen and compared, the extensive was chosen for further analysis. This choice is justified by the following X reasons: First, when comparing how conclusive composite task measures were filled, no particular RTI excelled over the other. While the manual CTMs were not filled to full satisfaction, no descriptive advantage of the aggregated RTI over the extensive RTI emerged. In terms of values, the aggregate RTI was limited in diversity through incorporating only two equally weighted CTMs, whereas the extensive RTI used five. As a result, individuals who scored particularly high or low in multiple CTMs, hence covering a diverse range of non-routine labour, enjoyed particular emphasis with the extensive RTI. Through its higher dimensionality and improved resolution, the extensive RTI more accurately reflects the contributions from a diverse range of non-routine labor aspects, giving a more rounded view of the RTI landscape. Moreover, the results of the extensive RTI were more conclusive, allowing for a more nuanced pattern of insights.

3.3. Suitability for Machine Learning

Suitability for Machine Learning (SML) describes susceptibility to automation by more sophisticated ML models. Beyond computerisation, as captured by RTI, SML also describes replacement potentials of tasks involving interpretation and generation of text and images, or more complex mathematical tasks, such as time-series predictions (Brynjolfsson, Mitchell, and Rock 2018). This includes generative AI, such as Large Language Models (LLMs), the technology behind Chat-GPT. Conceptually, SML follows the task-based labour market model as proposed by Autor, Levy, and Murnane (2003), and was devised to propose a partially quantified framework to the projection of Levy (2018), that ML could theoretically but will not immediately replace tasks as soon as technologically available. The SML-index consequentially only covers the technological potential of a task to be replaced; it's suitability. It does not give clear indication on whether a task will be automated soon. After all, such a projection would require a robust model depicting the complex interplay between institutional, management and economical factors around AI diffusion to a level of precision sufficient for prediction; a rather daunting imagination. Rather, SML is to be understood as the "upper boundary" of what can be automated: You can state that a task which is not SML cannot be automated through ML (Brynjolfsson, Mitchell, and Rock 2018).

The theoretical framework behind the index was developed earlier in a companion paper published in *Science* (Brynjolfsson and Mitchell 2017), where the authors identified eight criteria for task-SML (Table 3.1), extended to a 21-item rubric.

Task	**Explanation**
Function Mapping	Suitable for tasks where a function can map well-defined inputs to well-defined outputs, such as classification and prediction. Examples include labeling images or analyzing loan applications for predicting default likelihood.
Availability of Large Digital Data Sets	ML algorithms perform better with large digital data sets. Data can be created by monitoring processes, customer interactions, human tagging, or simulations.
Clear Feedback and Goals	Tasks that provide clear feedback with definable goals and metrics are suitable. ML works well when goals are describable even if the process to achieve them is not defined.
Short Logical Chains	ML excels in tasks requiring quick reactions based on empirical associations in data but struggles with tasks requiring long chains of reasoning or understanding of complex, unknown background knowledge.
Limited Explanation Requirement	Tasks where detailed explanation for the decisions made is not a priority. Despite ongoing research, current ML systems, especially deep neural networks, find it challenging to articulate the reasoning behind their decisions in human-understandable terms.
Error Tolerance	Suitable for tasks where there is tolerance for error and no need for provably correct or optimal solutions, as ML algorithms derive solutions statistically and probabilistically.
Stable Phenomenon	Tasks where the phenomenon or function doesn't change rapidly over time. ML requires that the distribution of future test examples is similar to the distribution of training examples.
No Specialized Physical Skills	Suitable for tasks that do not require specialized dexterity, physical skills, or mobility, as robots are still relatively clumsy compared to humans in physical manipulation.

Table 3.1: Most Suitable Tasks for Automation, source: Brynjolfsson and Mitchell 2017

The SML-index is a result of the follow-up study performed by Brynjolfsson, Mitchell, and Rock (2018), in which said rubric-items were applied to all O*NET DWA task-descriptions using Crowdflower, a ”human intelligence task crowdsourcing platform” (Brynjolfsson, Mitchell, and Rock 2018). The results revealed new occupations to be at the potential forefront of automation, such as brokerage clerks or credit authorisers. On the low end of the SML-distribution occupations involving manual labour, such as plasterers or massage therapists, or navigation in unstructured environments, like archeologists, were to be found, pairing well with the AI-bottlenecks as described by Frey and Osborne (2013).

Another relevant measure concerning SML is its standard deviation within occupations, sdSML. A high value for sdSML indicates the bundling of tasks with high as well as low SML. As a share of task will be automated, other tasks can’t however, it is likely that occupations with high sdSML will develop towards more specialised occupations focusing mainly on non-automatable tasks, congruent with the hypothesis of transformation. The correlation between SML, just as sdSML, with occupational wages has been found to be very low, indicating occupations subject to future automation might differ from the prior ”hollowing out”-pattern. Going beyond technical constraints of automation, it seems plausible however that especially high-wage occupations were to be targeted first, as they impose the highest financial incentive to automate. (Brynjolfsson, Mitchell, and Rock 2018)

3.3.1. Transferral of Task-SML to STEP

The transferral to STEP is inhibited through the fact that SML-scores are available for DWAs in O*NET, however SML for tasks within STEP remains unknown. Alternative to a crowdsourcing-based procedure, as performed Brynjolfsson, Mitchell, and Rock (2018), this thesis makes use of the technology in spotlight, and instead uses a modified version of Google’s language model (LM) BERT, SBERT (sentence-BERT, Reimers and Gurevych 2019), to assess the similarity between task items in STEP and DWA task descriptions from O*NET. This method has been inspired by Carbonero, Davies, et al. (2023), and refined in minor fashion. There are three direct benefits to the methodology of Carbonero, Davies, et al. (2023): First, it provides a resource-efficient solution to the assessment of SML in developing economies. It outperforms both crowdsourcing and assessment of similarity by a human in its requirements of time as well as monetary resources. Second, as the SBERT-LM proceeds inputs in a consistent fashion, reducing author bias compared to manual selection of ”equivalents” to STEP tasks in a set of 2074 DWAs. Unlike humans, an LM adheres to the same rules irrespective of a datasets size. Third, SBERT contains a module that is specifically trained in quantifying sentence similarity (Reimers and Gurevych 2019), unlike survey respondents or the author(s) of studies on automation.

In a first step, a similarity matrix between STEP task items and DWAs is created. This matrix contains cosine similarities, in which a perfect overlap is allocated the value 1, and a polar opposite would be awarded -1. Carbonero, Davies, et al. (2023) then calculate task-SML in STEP by using a weighted average of DWA-SML, weighted by similarity (Equation 3.3).

$$SML_{task} = \sum_{DWAs} \frac{similarity_{DWA,task} * SML_{DWA}}{\sum\limits_{DWAs} similarity_{DWA,task}} \tag{3.3}$$

This thesis deviates slightly in the calculation of SML_{task}. Whereas Carbonero, Davies, et al. (2023) sum over all 2074 DWAs, we only sum over a selected set. This is motivated by the reasons. First, it is deemed unfeasible to interpret the meaning of a negative value: While certain actions do have polar opposites such as *give* and *take*, it is unclear what polar opposites of work items such as *Read forms* would be, or whether they exist at all. Second, the large number of DWAs raises concerns about noise within the data. After all, eleven DWAs with a similarity of 0.1 would outweigh one DWA which is a perfect match to a STEP task. Consequently, few DWAs with high similarity are deemed as more conclusive than many with little or negative similarity. Third, like RTI, SML serves to describe differences in automation potential within a set of tasks and thus derives part of its meaningfulness from the deviation to the mean; its standard deviation. As per the nature of weighted averaging, especially when the weights have upper and lower boundaries, the standard deviation decreases with the increase of weighted factors, nearing all values to the mean-SML. This also suggests that a reduction in similar DWAs to the most similar ones is preferable over an uncurated set of DWAs.

Moreover, two contingencies were identified which might reduce methodological validity. It is possible that for a STEP-task some DWA task descriptions have high linguistic similarity, but are meant within a different context. This originates from the data structure within O*NET, where work activities are layered three-fold: General Work Activities (GWA), Intermediate Work Activities (IWA) and DWAs, of which the former two define the context, but were not assessed for SML (Brynjolfsson, Mitchell, and Rock (2018)). Additionally, it could occur that no significantly similar DWAs are found for a STEP task.

The considerations in weighting and potential impediments of validity motivate the following deviations from Carbonero, Davies, et al. (2023): Instead of averaging over the total set of DWAs, the similarity matrix is first refined further. For each STEP task, a candidate set of similar DWAs is attained by searching for the 5 most similar DWAs, all other similarities – or weights – are set to zero. This effectively reduced the reduction in standard deviation, reduced noise and filtered out weights lower than zero. Further, the remaining dataset was analysed manually for STEP tasks without significantly similar DWA-counterparts, of which fortunately none were found, and linguistically similar but contextually different DWAs were sorted out. Admittedly, the number 5 here is picked arbitrarily, however is large enough to reliably provide a set of similar task descriptions while being small enough to not introducing the aforementioned issues. The resulting similarity matrix can be found in the supplementary materials of this thesis.

3.3.2. Further Procedure

After calculating task-SML for STEP, the methodologies of Brynjolfsson, Mitchell, and Rock (2018) and Carbonero, Davies, et al. (2023) differ in level of aggregation, but conceptually overlap. For O*NET, occupations are directly described at hand of their activities, whereas STEP observations are per individual, requiring individual SML-aggregation first.

$$SML_i = \sum_{tasks} \frac{usage_{task} * SML_{task}}{\sum\limits_{tasks} usage_{task}} \tag{3.4}$$

Finally, to obtain the occupation level SML, a final weighted average is calculated, this time weighted by the representativeness weight provided by the STEP survey.

3.4. Socio-Demographic Cluster Analysis

The results from 3.2 and 3.3 explain how susceptible individuals and occupations are to having parts of their task composition automated. As reflected in the motivation of this thesis' second research question, the direct societal implications behind occupational transformations however require a more comprehensive analysis that extends beyond mere technological susceptibility. In particular, potentially uneven impacts of automation across various demographic groups calls for a nuanced understanding that can only be achieved through socio-demographic clustering. This approach is particularly crucial for two primary reasons: First, the impact of automation is not uniformly distributed across society; it varies based on factors such as age, education level, and gender, and income. Recognising these disparities allows for the development of targeted policy interventions that are both economically efficient and socially equitable. Second, in the context of emerging economies, which are characterised by higher levels of informal labour, lower educational attainment, and weaker social safety nets, this level of analysis becomes even more crucial. In such circumstances, automation's social effects may be severe and immediate, possibly escalating already-existing social inequalities. Understanding the unique vulnerabilities of various demographic groups becomes crucial given the constrained institutional resources and capacities of emerging economies.

In this context, socio-demographic clustering provides an initial separation of socio-demographic groups. We can identify trends and possible discrepancies in how automation affects different layers of society by dividing the workforce into multiple groups according to criteria like education, age, financial situation, or employment type. People in unstable jobs or those with low levels of education, for instance, could find it more difficult to move into new careers than people with in-demand skill sets. Similar to this, the automation of highly prevalent professions might result in widespread unemployment, putting high pressure on social welfare systems. Socio-demographic clustering enables a more granular perspective on society behind the labour market and facilitates to address specific needs and challenges faced by different demographic cohorts, thereby fostering a more inclusive and sustainable policy net against automation. Table 3.2 depicts the chosen dimension of clustering, and motivates the choice in the light of this thesis' analysis

3.4.1. Technical Motivation for Clustering

Clustering is a form of unsupervised machine learning that involves grouping data points into subsets (clusters) such that data points in the same subset are similar to each other, but different to other clusters. The underlying dynamic to facilitate this is an optimisation interplay between minimisation problems (cluster-internal) and a meta-scale maximisation problem (cluster-overarching) (MacQueen 1967, Bishop 2006). Taking into account the fact that our data is based on individuals of developing countries, which's societal makeup lies out of the author's expertise, unsupervised ML algorithms like most clustering algorithms, are particularly suitable as they require no pre-labeled data.

K-means clustering is one of the most used clustering methods, mainly motivated in its simplicity and effectiveness. Each data point is assigned to the cluster with the closest mean after the dataset is divided into k clusters. The technique minimises the sum of squared distances from each point to its assigned centroid by repeatedly adjusting the placement of the centroids (cluster centers). Due to the tendency of being computationally cheaper than other clustering methods, K-means is particularly well suited for datasets with a relatively small number of variables, as in our case. (Bishop 2006)

The choice of k-means clustering is rooted in three reasons: First, thanks to libraries like scikit-

Dimension	Motivation
Education Level	Those with higher education levels may have more transferable skills, easing transitions to new roles. Assessing vulnerability of education groups can help target re-skilling and up-skilling programs.
Age Group	Younger individuals may be more adaptable and or receptive to acquiring new skills, especially given this would affect a larger share of their total career. Age-tailored strategies are imperative for the decision between vocational restructuring or strengthening of social nets.
Gender	Social norms on gender often play a role in employment form and what jobs and tasks are carried out. Gender differences in vocational changes can help in the evaluation of parity policies.
Income Level	Members of lower-income groups might face higher barriers when looking for new occupations. Understanding the economic impact aids the development of financial support programs.
Employment Status	Self-employed workers might have different challenges compared to full-time employees. Understanding these differences can help in developing tailored support programs.

Table 3.2: Dimensions and Motivations for Socio-Demographic Clustering

learn, which offers a streamlined version of the k-means method, Python offers a strong environment for for k-means clustering. Second, k-means suits this thesis given its effectivity in handling datasets with few variables. And last, k-means is one of the most broadly and frequently used methods, rendering its results reliable without requiring too high resource-allocation to validation (MacQueen 1967, Bishop 2006).

3.4.2. Clustering with Mixed Data

Clustering in its simplest form is performed at a set of variables of the same kind. Its nature of an unsupervised distance-optimising algorithm renders it to be applied the easiest on normalised continuous variables. However, in this thesis' research, categorical variables like educational attainment, gender or employment status, impose challenges on the underlying optimisation algorithm. If not treated with special care, steps in binary variables constitute a direct jump between 0 and 1, the maximal distance on a normalised scale, and outweigh all non-maximal differences in continuous dimensions. As a result, the clusters will be generated first at hand of all binary variables, then low-granularity categorical variables, and only then continuous and high-granularity categorical variables.

One potential avenue of addressing this issue is the choice of alternative algorithms which allow for the incorporation of multiple different data types. In the realm of clustering algorithms, potential alternatives to K-means would be Spectral Clustering, Agglomerative Clustering, and Gaussian Mixture Models (GMM). Spectral Clustering is particularly beneficial at complex, non-spherical cluster structures but is computationally demanding and requires a priori specification of cluster numbers (Luxburg 2007). Agglomerative Clustering offers a hierarchical approach that does not necessitate pre-analysis definition of cluster numbers, but encounters difficulty in classifying data points in large datasets, particularly with many observations (Murtagh and Legendre 2014)). Gaussian Mixture Models provide flexibility in cluster covariance and can model elliptical clusters, but are prone to overfitting and require high preparatory effort (R. 2009).

An alternative approach to tackle variable imbalances would be the manual adjustment of weights of binary and categorical variables. This approach is more straightforward, but tackles the imbalance of dimensions effectively. The adjustment of weights would take place in the reduction of ranges, or differences in values in between binary or categorical variables. If the differences between two values of a binary variable resembles the standard variation in other variation, the variables are weighted similarly in the minimisation and maximisation of differences. In essence, this is comparable to how many mixed-data clustering methods deal with different data types (like DBScan), but less opaque and open to refine in an iterative fashion. Through this iterability, clusters can eventually be improved in quality.

Upon comparing the alternative algorithms and the option of manual weight adjustment, the latter was chosen for three reasons. First, iterative weight adjustment provides higher control and transparency over the factors that the k-means emphasises than the choice of other algorithms would. Second, it is more scalable because it doesn't have the computational requirements of some alternative algorithms. Lastly, it permits the application of the original clustering algorithm, which is significantly more established for socio-demographic analyses success in solving the unique problems presented by socio-demographic data.

3.5. Multivariate Linear Regression Models

To determine whether the effects observed for individual occupations, clusters, or socio-demographic traits and their respective automation potentials are of statistical significance, and to assess impact size, multivariate linear regression models are used. Multivariate linear regression models provide information on how the change of a single variable, as all other variables remain unchanged, causes effect on the goal variable: our automation indicators. This mainly serves to validate descriptive observations in the results section. Per indicator and per country, three regression models will be built. First, occupational influences will be considered. For this, dummary variables for the one-digit ISCO-codes, providing high-level occupational information will be compared towards each other (Equation 3.5 and Equation 3.6).

$$RTI = \sum_{i=1}^{9} \beta_i I_{ISCO=i} \tag{3.5}$$

$$SML = \sum_{i=1}^{9} \beta_i I_{ISCO=i} \tag{3.6}$$

Where the β coefficients represent the linear regression coefficients and thus impact sizes of the dummy variables $I_{ISCO=i}$ which per digit i represent the affilitation of an observation to the respective one-digit ISCO code. Second, socio-demographic characteristics sex, age, education and income will be considered. Income will be considered unorthodoxly, as we are dealing with many different currencies and purchasing power parities (PPP). While it is possible to transform national currency incomes into US-dollar PPPs, this provides us with little indication of how an individual performs within the national income distribution. Alternatively, the standard deviation from the national mean of logarithmised income is used. This provides us human-interpretable insights on what one unit of change in income means in relative terms to the income distribution (Equation 3.7 and Equation 3.8).

$$RTI = \beta_1 Female + \beta_2 Age + \beta_3 LowEd + \beta_4 HighEd + \beta_5 \left(\frac{\ln(Income) - \mu_{Income}}{\sigma_{Income}} \right) \tag{3.7}$$

$$SML = \beta_1 Female + \beta_2 Age + \beta_3 LowEd + \beta_4 HighEd + \beta_5 \left(\frac{\ln(Income) - \mu_{Income}}{\sigma_{Income}} \right) \tag{3.8}$$

Where $Female$, Age, and $Income$ are self-descriptive, and $LowEd$, and $HighEd$ are dummy variables representing an education level either below or above high-school level. μ_{Income} represents the mean and σ_{Income} represents the standard deviation of $Income$ in the respective country. Third, influences of cluster affiliation on automation indicators will be analysed (Equation 3.9 and Equation 3.10).

$$RTI = \sum_{i=1}^{n} \beta_i I_{Cluster=i} \tag{3.9}$$

$$SML = \sum_{i=1}^{n} \beta_i I_{Cluster=i} \tag{3.10}$$

Where the dummy variables $I_{Cluster=i}$ represent affiliation to cluster i. Traditionally, multivariate regression models also are provided a constant, which provides a baseline value from which the independent variables exert influence upon the dependent variable. In this thesis however, this is not required as our automation indicators are standardised, which renders their mean value zero. The omission of constants in fact allows for a better representation of ceteris paribus effects.

Results

In this section, the results of our analysis are presented. It is important to note that the socio-demographic cluster analysis is performed on a country-only basis. Yet, to provide an initial, high-level, sense of how RTI and SML are associated to occupations and the socio-demographic clustering variables, the first subsection describes these relationships on a global level. The following subsections contain the analysis of SML and RTI against socio-demographic indicators as well as the results of the cluster analysis on a country-level. As the country-level analysis in it is quite repetitive, only partially raising insights of novel nature with every new country, not all countries results will be presented in full detail. Only the results for the first country, Colombia, will be presented in full detail. The sections for the countries following Colombia will mainly discuss insights unique to the respective country, and only a brief listing of replicated patterns. Colombia as the first and only country to be considered fully is admittedly chosen arbitrary, based more on Colombia serving as the prototype for analysis than on any added informative value unique to Colombia. The visualisations which accordingly are left out to prevent repetition are to be found in Appendix D. The chapter ends with an interim summary.

4.1. Global

This subsection present the results of our analysis on a global scale, meaning for all countries of analysis. This first includes the occupational results per indicator, followed by the distribution of RTI and SML over socio-demographic variables. For both occupations (1 digit ISCO-codes) and socio-demogrpahic variables, the results of the first two regression models out of Section 3.5 are provided. Additionally, the occupations with the highest and lowest values for SML and RTI are listed within the respective chapters. The discussion of clusters and influences of wage are kept on a country specific-level due to expected issues in comparability on a global scale. Finally, comparisons to findings of prior research for developed economies are made.

4.1.1. Routine Task Intensity

Occupational Results

Table 4.1 depicts the regression coefficients of ISCO-classification on RTI over our countries of analysis (A link between ISCO-codes and the occupations and skill-levels they represent is to be found in Appendix A). These coefficients indicate the change in the Routine Task Intensity (RTI) index corresponding to a one-unit change in each dummy variable, while holding other variables constant. From ISCO 1 to ISCO 4, the pattern is a robustly negative coefficient of occupational affilitation on RTI. ISCO 1 to ISCO 3 resemble high-skill occupations, which on average show the lowest RTIs. ISCO 4, clerical service workers and thus characteristically most customer interaction, also show lower RTI levels. With increase in digit this trend reverses. For ISCO 5, service and sales workers, the coefficients are comparably small and alternate in polarity. With the exception of ISCO 8, the high-digit occupations, those with lower skill-level, robustly show positive RTI scores. The highest RTIs are found for agricultural and fishery workers (ISCO 6) and elementary workers (ISCO 9).

	Columbia	Sri Lanka	Laos	Vietnam	Yunnan	Global
$\beta_{ISCO=1}$	-0.35***	-0.52***	-1.10***	-0.31***	-0.25***	0.03***
$\beta_{ISCO=2}$	-0.75***	-0.45***	-0.74***	-0.67***	-0.34***	-0.60***
$\beta_{ISCO=3}$	-0.35***	-0.74***	-0.66***	-0.52***	-0.29***	-0.34***
$\beta_{ISCO=4}$	-0.39***	-0.72***	-0.94***	-0.18***	-0.10***	-0.48***
$\beta_{ISCO=5}$	0.00	-0.11***	-0.11***	0.12***	0.12***	-0.14***
$\beta_{ISCO=6}$	nan	0.54***	0.44***	0.31***	0.36***	0.43***
$\beta_{ISCO=7}$	0.25***	0.21***	-0.21***	0.58***	0.45***	0.47***
$\beta_{ISCO=8}$	-0.12***	-0.85***	-1.41***	0.65***	-0.04	0.55***
$\beta_{ISCO=9}$	0.45***	0.54***	0.46***	0.05***	0.32***	0.04***

Table 4.1: Regression Coefficient of first-digit ISCO on RTI per country (***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$, No asterix: No significant association)

Table 4.2 depicts the occupations with the highest and lowest RTIs on a two-digit ISCO level (minimum 10 observations per occupation). Comparing the top and bottom RTI occupations, a contrast in educational level emerges. While none of the top-RTI occupations requires university level-education, the low RTI occupations (except of authors) all require a university degree. The more granular perspective on occupations also allows for a quality examination of our calculated index; a comparison between what has been calculated and what aligns with the definition of routine by Autor, Levy, and Murnane (2003). For the low-RTI occupations, it becomes apparent that our index covers occupations which have a high share of non-routine tasks very well. For high RTI occupations however, it must be noted that truck and bus drivers are not aligned with what is understood as routine by Autor, Levy, and Murnane (2003). Undoubtedly, the responsibilities of truck and bus drivers exceed the task of driving, yet is driving their main activity, in time and effort. Bus and truck drivers are a good example for how our index to some extent still falls short to measure task importance as described in Section 3.2.2.

High RTI Occupations	
Occupation	**RTI**
Locomotive engine drivers and related workers	1.007
Painters, building structure cleaners and related workers	0.981
Wood treaters, cabinet-makers and related trade workers	0.959
Heavy truck and bus drivers	0.906
Blacksmiths, toolmakers and related trade workers	0.868
Low RTI Occupations	
Legal professionals	-1.477
Software and applications developers and analysts	-1.406
University and higher education teachers	-1.308
Authors, journalists and linguists	-1.190
Other teaching professionals	-1.088

Table 4.2: Comparison of High and Low Routine Task Intensity (RTI) Occupations

Socio-Demographic Results

A few noteworthy observations can be taken from the left part of Figure 4.1: RTI distribution differs from in form and median with respect to gender. Globally, men generally show higher values in RTI than women and show a bulk of the statistical population in between the half and one and a half standard deviations above the global mean, with little observations exceeding one and a half standard deviations. The distribution of RTI across various age groups is seen in the right plot of Figure 4.1. We observe that the distribution for the youngest workers seems more shifted towards the extremes than for other age groups with spikes on the left and mid-right end of the distribution. For the higher age groups the distribution is more close to normal. It can also be observed that the median RTI initially increases over age, then drops again for the highest age group.

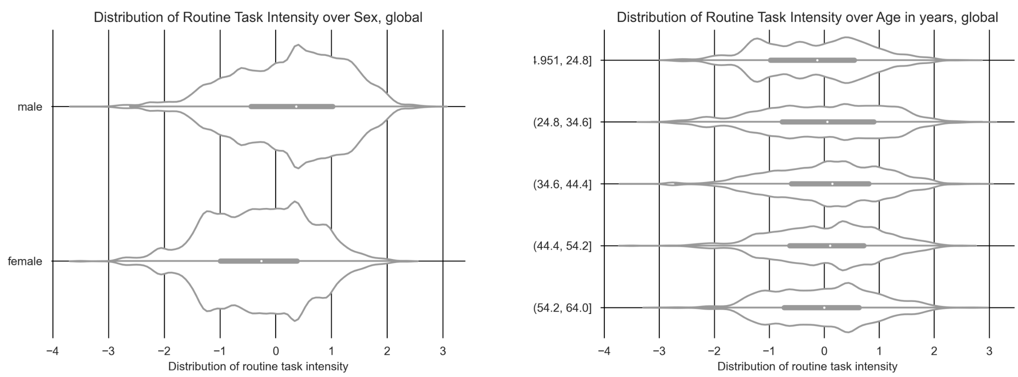


Figure 4.1: Global RTI Distribution against Sex and Age (source: (World Bank 2014), own calculations)

Figure 4.2 depicts the distribution of RTI over level of education, where low-education refers to no attainment of a secondary school degree or attendance of a secondary school, mid-education refers to the attainment of a second-degree school degree and high-education describes those who hold at least a bachelor's degree or equivalent. A clear tendency is seen over increasing educational attainment: RTI decreases. This contrasts with the findings on developed economies, where high RTI was characteristic to middle-education occupations, not low-education (Autor, Levy, and Murnane 2003, Autor and Dorn 2009, Goos, Manning, and Salomons 2014). One potential explanation of the high RTI associated with lower education levels might be understood in the context of greater competitiveness around educational attainment. Given the comparatively lower availability of secondary education degrees in emerging economies, having one can provide an advantage in the labour market, leading to more advanced posts with lower RTI. This is in contrast to the situation in industrialised economies, where a secondary degree is more widespread and often leads to typical office employment with greater RTI. As a result, in emerging economies, the higher competitive value of a secondary degree can often lead to people shifting away from routine-intensive employment and towards more complex roles.

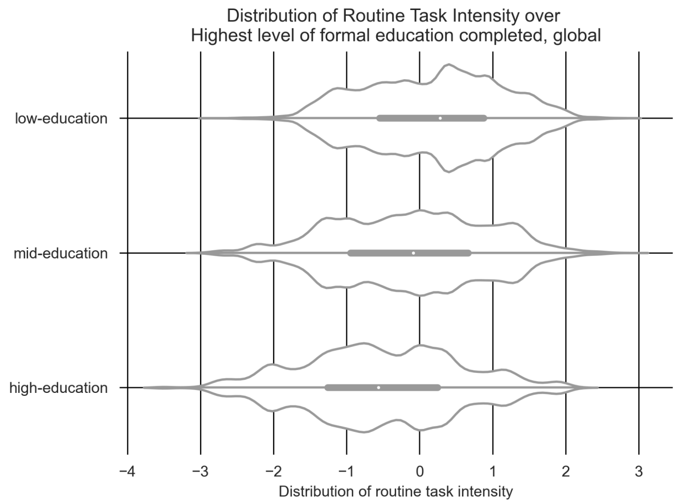


Figure 4.2: RTI Distribution against Sex and Age for Colombia (source: (World Bank 2014), own calculations)

Below, Table 4.3 illustrates the linear regression coefficients of various control variables, which are listed in the first column. Each row represents a different variable, including a dummy variable female, age, income, a dummy variable for low Education, and dummy variable high education. The coefficients for these variables across different geographical locations (Columbia, Sri Lanka, Vietnam, and Yunnan) as well as the global aggregate are presented from the second to the sixth columns. These coefficients indicate the change in the Routine Task Intensity (RTI) index corresponding to a one-unit change in each control variable, while holding other variables constant.

	Columbia	Sri Lanka	Vietnam	Yunnan	global
β_{Female}	-0.09^{***}	0.13^{***}	-0.21^{***}	0.07^{***}	-0.29^{***}
β_{Age}	0.01^{***}	0.00^{***}	0.01^{***}	-0.00^{***}	0.01^{***}
β_{Income}	-0.54^{***}	-2.26^{***}	0.24^{***}	-0.56^{***}	0.14^{***}
β_{LowEd}	0.00	0.11^{***}	0.16^{***}	0.07^{***}	0.20^{***}
β_{HighEd}	-0.37^{***}	-0.42^{***}	-0.59^{***}	-0.37^{***}	-0.43^{***}

Table 4.3: Socio-Demographic Regression Results for RTI (***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$, No asterix: No significant association)

Note that some variables have alternating polarities for different countries, especially sex. With the exception of Vietnam, the highest association to RTI is mostly seen for income, with consistently negative coefficients. Education and RTI also have a consistently negative relationship. This shows a robust global pattern of increases in educational attainment and income being predictive of a decrease in RTI with high statistical confidence ($p < 0.01$ for all countries).

4.1.2. Suitability for Machine Learning

Occupational Results

Table 4.4 presents the regression coefficients of affiliation to ISCO occupational codes on SML. Contrary to RTI, the high-skill occupations in ISCO 1 to ISCO 3 and the clerical workers in ISCO 4 show the highest values for SML. ISCO 5 shows robust positive values and hence above average automation potentials. Elementary workers (ISCO 9) tasks are the least SML, followed by plants and machine operators, who are robustly below average SML (ISCO 8).

	Columbia	Sri Lanka	Laos	Vietnam	Yunnan	Global
$\beta_{ISCO=1}$	0.44^{***}	0.50^{***}	0.52^{***}	0.29^{***}	0.17^{***}	0.46^{***}
$\beta_{ISCO=2}$	0.20^{***}	0.64^{***}	0.78^{***}	0.23^{***}	0.11^{***}	0.44^{***}
$\beta_{ISCO=3}$	0.36^{***}	0.39^{***}	0.53^{***}	0.15^{***}	0.30^{***}	0.39^{***}
$\beta_{ISCO=4}$	0.38^{***}	0.47^{***}	0.54^{***}	0.38^{***}	0.28^{***}	0.43^{***}
$\beta_{ISCO=5}$	0.05^{***}	0.37^{***}	0.29^{***}	0.06^{***}	0.03^{**}	0.17^{***}
$\beta_{ISCO=6}$	nan	-0.20^{***}	-0.29^{***}	0.09	-0.91^{***}	-0.39^{***}
$\beta_{ISCO=7}$	0.16^{***}	0.06^{***}	0.07^{**}	-0.12^{***}	-0.49^{***}	0.01^{*}
$\beta_{ISCO=8}$	-0.18^{***}	-0.38^{***}	-0.51^{***}	-0.47^{***}	-0.29^{***}	-0.37^{***}
$\beta_{ISCO=9}$	-0.61^{***}	-0.60^{***}	-0.37^{***}	-0.75^{***}	-0.36^{***}	-0.66^{***}

Table 4.4: Regression Coefficient of first-digit ISCO on SML per Country (***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$, No asterix: No significant association)

Table 4.5 shows the most and least SML occupations. Among high SML, occupations of high and low specialisation, or education, can be found, which overall provides a more nuanced image of what occupations are SML in particular when compared to RTI. With the exception of fishery workers and hunters, the five least SML occupations all are elementary occupations, with low educational requirements and pay. These roles predominantly fall under manual or labor-intensive categories, indicating that they are less susceptible to automation and machine learning technologies. However, the presence of diverse roles in the high SML category, such as cashiers and ticket clerks, adds complexity to this pattern. This diversity suggests that the relationship between job roles and their susceptibility to machine learning is not as straightforward and warrants further nuanced investigation.

High SML Occupations	
Occupation	**SML**
Process control technicians	0.935
Database and network professionals	0.851
Nursing and midwifery associate professionals	0.723
Retail and wholesale trade managers	0.716
Cashiers and ticket clerks	0.695
Low SML Occupations	
Fishery workers, hunters and trappers	-1.428
Assemblers	-1.358
Domestic, hotel and office cleaners and helpers	-1.163
Vehicle, window, laundry and other hand cleaning workers	-0.976
Locomotive engine drivers and related workers	-0.876

Table 4.5: Comparison of High and Low SML Occupations

Socio Demographic Results

Figure 4.3 reveals a left skew in the distribution of SML over both genders and all age groups, with medians above the global mean and a tail of observations exceeding over up to three Z-values to the left. This skew suggests two insights: For Machine Learning, we expect a small share of ”resilient” occupations, likely those which primarily comprise ”engineering bottlenecks” in their task composition (Frey and Osborne 2013). Secondly, the skew suggests that the majority of workers has above average SML. This is to interpret with caution however and does mean they are particularly exposed to occupation, as we only observe standardised SML, no more than deviations from the mean, instead of absolute an SML. Yet, we can deduce a potential of unequal automation from this skew, given that few occupations might be underproportionately suited for automation, whereas the bulk seems to be affected similarly.

The left part of Figure 4.3 shows the distribution of SML over gender, with the majority of observations seeming to be more normally distributed than for RTI. Women show a higher median SML and larger proportion of high SML ($Z > 1$) than men, suggesting their tasks are generally more SML as compared to men. It is important to note that the left tail – creating the left skew and resembling those who are only marginally affected by SML – is observably thicker for men. This implies that the occupations ”resilient” to ML are particularly male in profile, which would be a clear deviation from earlier automating technologies represented in RTI.

The right side of Figure 4.3 shows the distributions of SML over differing age groups. When considering the shifts towards higher or lower tendency in SML, a noteworthy observation emerges: The shifts resemble those across RTI (Figure 4.1) quite closely. We see the lowest median for the youngest and oldest age group, with the prime-age workers having higher median values, to be observed the clearest for workers between 35 and 44. For SML however we note that the median SML for the younger age group between 25 and 34 also is closely located to the maximum median, rendering SML a ”younger” phenomenon than RTI, which might be a result of change in skill demand for the more recent entrants of the labour market.

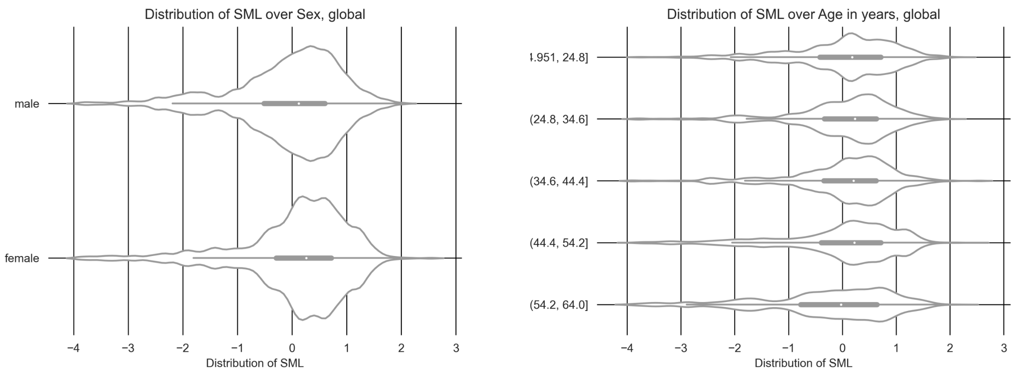


Figure 4.3: Global SML Distribution against control variables (source: (World Bank 2014), own calculations)

Figure 4.4 shows the distribution of SML over low, middle, and high-education. It can be observed that with an increase of educational attainment, the distribution shifts right, with a consistent increase in median value over the educational levels. It also becomes apparent that the spread decreases over increasing educational attainment: While the spread between the first of third quartile boundary spread over more than a whole standard deviation for the low-educated, the spread for the highly-educated is only a bit more than half of one standard variation, placing more than half of highly educated workers in the interval between 0.2 and 0.7 SML. Figure 4.9 also reveals that most of the left tail which skews the distribution, is comprised of individuals with low or no education. This suggests that the suspected ”resilient” occupations, might in fact require less skills acquired in formal education.

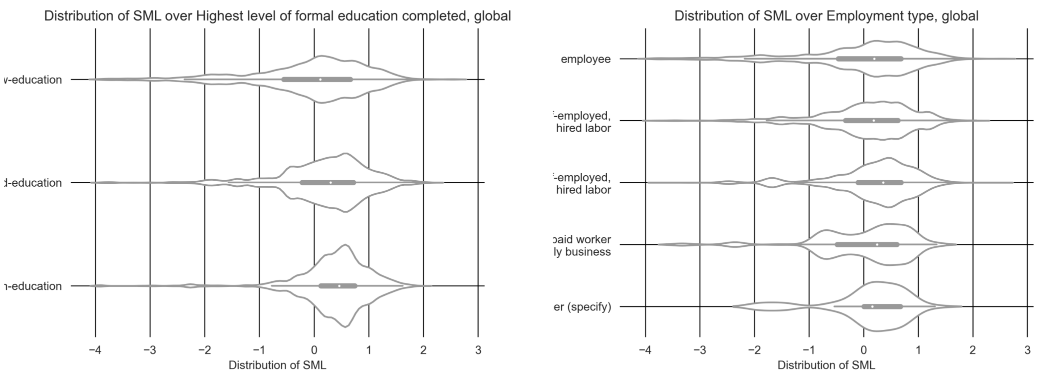


Figure 4.4: Global SML Distribution against control variables (source: (World Bank 2014), own calculations)

Below, Table 4.6 presents the regression coefficients β_{variable} for the socio-demographic clustering variables in the second to fifth columns' linear regression coefficients for the control variables for each country of analysis. The four columns on the right show the corresponding p-values p_{variable} for the coefficients for the various nations. Contrary to RTI, we see a more consistent re-

	Columbia	Sri Lanka	Vietnam	Yunnan	global
β_{Female}	0.04***	0.33***	0.08***	0.08***	0.14***
β_{Age}	0.00***	-0.01***	-0.00***	-0.00	-0.00***
β_{Income}	0.08***	0.58***	0.33***	-0.14***	0.29***
β_{LowEd}	-0.09***	-0.05*** *	0.20***	0.00	-0.02***
β_{HighEd}	0.22***	0.45***	0.18***	0.12***	0.28***

Table 4.6: Socio-Demographic Regression Results for SML (***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$, No asterix: No significant association)

lationship between sex and SML, which has no strong but consistently significant effect on SML. Income typically shows the largest correlation with SML and is mostly positive. However, there are deviations, for instance Laos ($p > 0.05$) and the Chinese province Yunnan ($\beta_{income} < 0$).

4.2. Colombia

This section demonstrates our results for the Colombian labour market. First, RTI and SML are discussed over the general statistical population in STEP. Hereafter the discovered socio-demographic clusters are presented and profiled, after which they are discussed for automation potential through digitisation and ML.

4.2.1. Routine Task Intensity

In the following, the distribution of RTI is shown over various socio-demographic control variables for the general statistical population. In particular, the distributions over sex, age, educational attainment and employment type are depicted and described, followed by the distribution of RTI over monthly income. The section ends with a description of high and low RTI-occupations within the Colombian labour market.

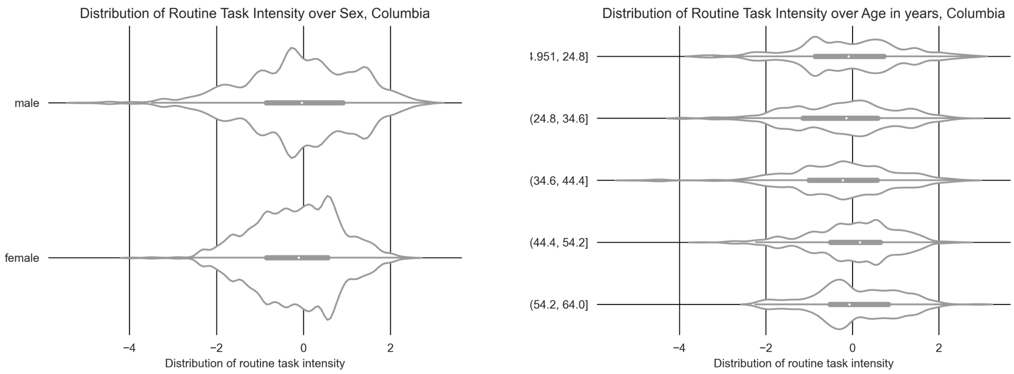


Figure 4.5: RTI Distribution against Sex and Age for Colombia (source: (World Bank 2014), own calculations)

A noteworthy observation emerging from the left part of Figure 4.5 is that men's RTI distribution is wider than women's. The interquartile range is wider for men even though the median RTI values for both sexes are close to identical, hovering around the global mean of zero. As a result, it appears that men hold a wider variety of employment with respect to RTI, including both low and high RTI roles. Women's jobs, on the other hand, are more centred around the median RTI value. Men's data show two distinct spikes, one smaller in the highest quartile and one below the median. Near the intersection of the third and fourth quartiles, there is a noticeable spike for women. These spikes indicate high concentration of jobs with RTI in the spiked areas, which is to be analysed later within the sociodemographic groups.

The distribution of RTI across various age groups is seen in the right plot of Figure 4.5. The interquartile ranges are noticeably wider for younger age groups' RTI values than for the two older age groups. This implies that younger people hold a wider range of job kinds, from low to high RTI tasks. The median RTI values show an intriguing pattern: they are lowest for the age group of 35 to 44 and noticeably high for that of 45 to 54. This may be due to a change in employment kinds as people get older, with older people likely working in more routine jobs with higher RTI values, such as outsourced service jobs, wehreas jobs with lower RTIs are typically taken by the characteristically highest-educated middle-age group.

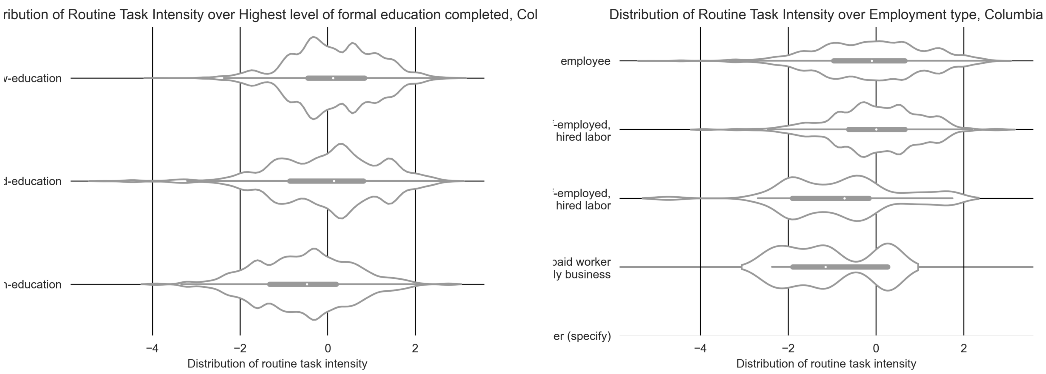


Figure 4.6: RTI Distribution against control variables for Colombia (source: (World Bank 2014), own calculations)

In a first glimpse, we see no clear tendency of RTI to be especially high for the mid-educated group in the left part of Figure 4.6, contrary to the findings of Autor, Levy, and Murnane (2003) for developed labour markets. While the median value of RTI is highest for the mid-educated group, and they show spikes in higher RTI ranges than the high- and low-educated groups do, the inter-quartile range for mid-educated workers is not visibly higher than for low-educated workers. Clearly apparent is the difference between high-educated workers towards other groups of lower education levels: Almost three quarters of the high-educated group is located below the global mean, and their median has a Z-value lower than -0.5, which implies significant and large differences in RTI with the attainment of higher education. One particular spike is visible close to two negative standard deviations, which potentially marks a social cluster with very little automation potential through digitisation. Among the variables controlled for, education seems to bear one of the largest and most significant correlations to RTI.

The right part of Figure 4.6 depicts the distribution of RTI over employment status. Self-employed workers without hired labour show no clear difference in RTI-tendency to employed workers except for a more narrow distribution around the mean. This could be due to the nature of self-employment without employees, which requires workers to take over all outstanding tasks, resulting in a very broad task set. A more clear difference can be seen between employers and self-employed or employees: They show lower values of RTI, with the median being close to a whole Z-value away of the global mean. A second spike can be seen two standard deviations away from the global mean, which implies digitisation often is less feasible than for other groups. This can be explained easiest through the delegation of clearly definable tasks to employees.

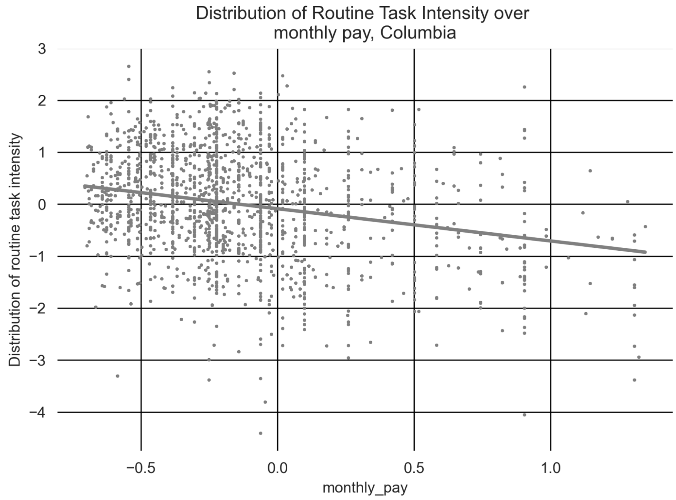
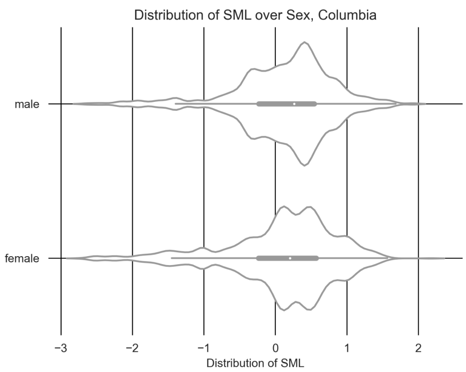


Figure 4.7: RTI Distribution against control variables for Colombia (source: (World Bank 2014), own calculations)

Figure 4.7 depicts the relationship between monthly pay and RTI in a regression plot. The emerging overall tendency is that we see a decline in RTI over increasing pay, which is in line with prior observations for groups typically associated to higher pay: high-educated workers or employers.

4.2.2. Suitability for Machine Learning

Figure 4.8 depicts the distribution of SML, the potential for task automation through sophisticated machine learning models, against the control variables Sex (left) and Age (right). On the left part, the SML distribution shows a right-skewed shape for both genders, with the median value being higher than the average. As can be seen on the left side of figure 6, this asymmetry points to the existence of a small subgroup of people with exceptionally low automation potential. The spread of SML values across occupations, in contrast to the RTI distribution, does not significantly differ between genders, suggesting that there is an equal distribution of employment across the range of machine learning applicability. The medians do not differ by gender either, indicating that men and women have nearly identical central tendencies for SML.



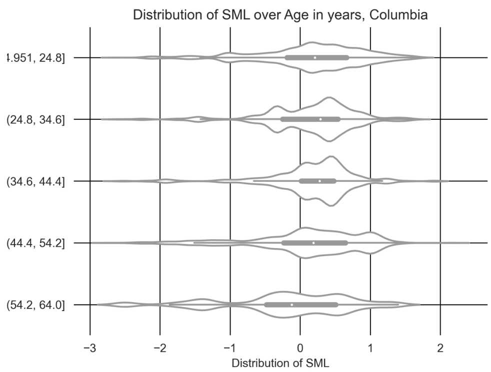


Figure 4.8: SML Distribution against control variables for Colombia (source: (World Bank 2014), own calculations)

On the right side of Figure 4.8, a pattern between SML and age can be observed. From young to old, there first is a minor increase in median-SML from the age range of 15–24 to 25–45, after which it enters a practically exponential drop. A more homogenous distribution of tasks that are SML may be indicated by the visibly lower spread of SML for the age group 35–44. Also we see an inverse pattern to RTI distribution, where the age groups between 25 and 44 demonstrated generally lower RTIs: For SML their medians are shifted the most to the right, for the age group 35-44 we see the three higher quartiles equal to or above the global mean.

Also contrary to RTI, the left part of Figure 4.9 depicts that typically groups with higher educational attainment also have higher susceptibility to ML automation. More so, the spread in distribution seems to decrease with higher educational attainment further increasing the likelihood of high SML for high-educated individuals. The right side of Figure 4.9 reveals that employers have lower SML and lower spread of SML than employees and self-employed workers without employees. This suggests that delegated tasks are more susceptible through ML-enabled automation than those taken up by business owners.

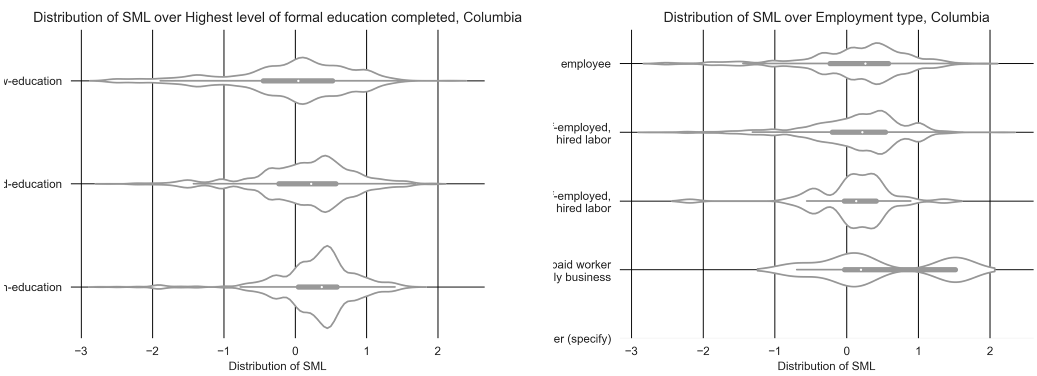


Figure 4.9: SML Distribution against control variables for Colombia (source: (World Bank 2014), own calculations)

Figure 4.10 depicts the relationship between monthly pay and SML in a regression plot. While overall a slight positive correlation between monthly pay and SML can be seen, it must also be noted that for individuals with high monthly pay the SML-values rather tend to converge slightly above 0, with few high values for SML which can be observed more often on the lower end of the distribution. One explanation for this could be that a linear regression might not be the best-suited curve here as one one hand, SML typically increases with education attainment, but on the other, business owners, which typically have the highest pay and are to be expected to the most right on the distribution, have lower SML. Overall, this suggests that at highest susceptibility to ML-automation we expect highly-educated employees, characteristically within the ages of 25-44 in Colombia with little to no influence of gender.

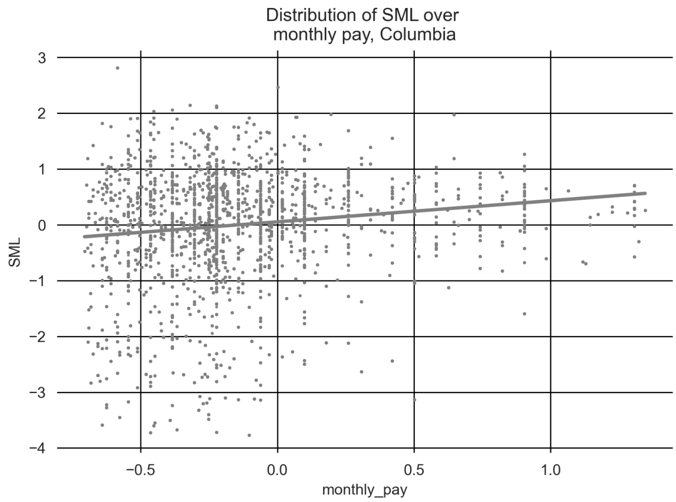


Figure 4.10: SML Distribution against control variables for Colombia (source: (World Bank 2014), own calculations)

4.2.3. Socio-Demographic Clustering

In this section, we first describe the clusters at hand of their socio-demographic characteristics, providing an initial picture on which parts of the labour force are represented within our clusters. Subsequently, SML and RTI is presented per cluster, facilitating insights in socio-demographic differences in automation potential between clusters. Finally, the results are compared to the general statistical population to outline unique patterns of socio-demographic clusters with respect to automation potentials.

Cluster Profiles

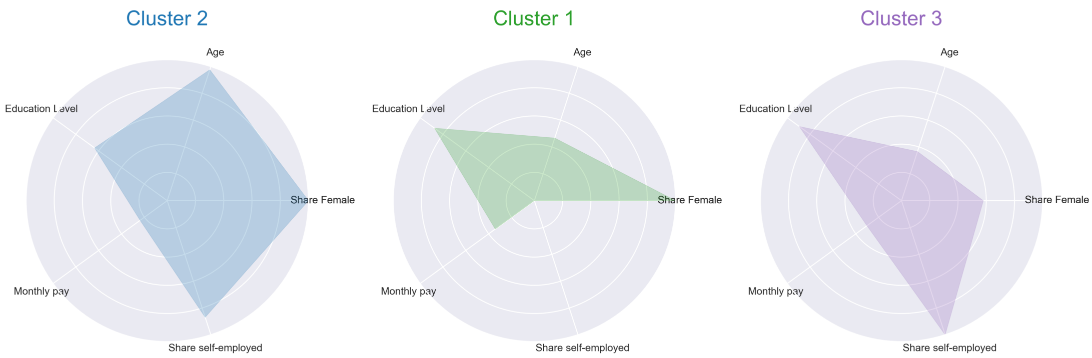


Figure 4.11: First Three Socio-Demographic Clusters Colombia (source: (World Bank 2014), own calculations)

Older, female mostly self-employed Workers with low educational attainment ***(Cluster 2)***: Females predominate in this cluster. This cluster is the oldest, with an average age of roughly 50. They have finished less formal schooling—on average, about 4 years, or primary education. They make less money on average each month, and 58 percent of them work self-employed. The most frequent occupations in this cluster are occupations in the garment industry, sales workers, or occupations in the hospitality sector.

Young working women with low-paying jobs ***(Cluster 1)***:
This cluster also is female-dominated and with an average age of 31 years, they are younger than other clusters and have completed a medium level of formal education (on average secondary education). None of them are self-employed, and their monthly salaries on average are low. They mostly work as employees. Frequent occupations in this cluster are shop saleswomen or other saleswomen, workers in the hospitality sector or client information workers.

Young, moderately-educated, freelancers ***(Cluster 3)***:
There are both males and females in this cluster, with the males having a slight edge. With an average age of 28 years and a moderate degree of formal education (on average approximately 5), they are the youngest category. They make less money on average each month, and 67 percent of them work for themselves. The workers in this cluster have no characteristical occupations, and overlap quite strongly with the other clusters. Their most frequent occupations are the generally most frequent occupations, with sales and elementary workers represented most, followed by shop salespersons. Accordingly, this cluster does not represent typical freelancer occupations in developed economies, like programmers or marketing professionals, but rather individuals partaking in the gig-economy.

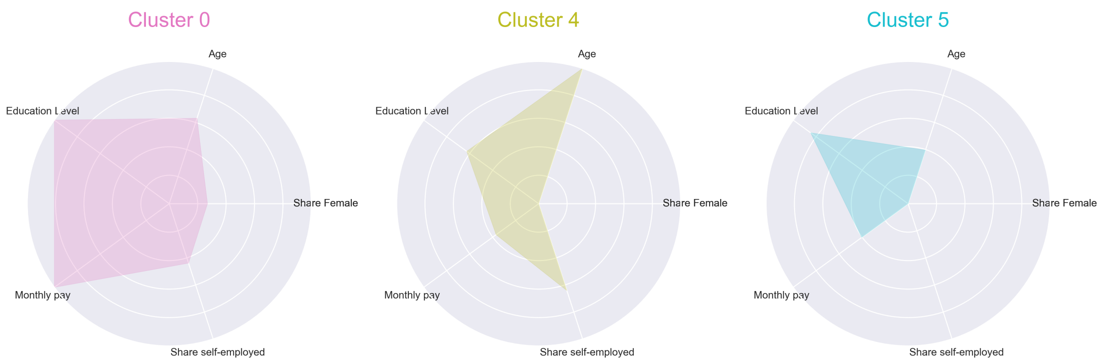


Figure 4.12: Second Three Socio-Demographic Clusters Colombia (source: (World Bank 2014), own calculations)

High-earning male professionals ***(Cluster 0)***:
Males predominate in this cluster. They have the highest degree of educational attainment (average score of 6, which corresponds to post-secondary education) and a relatively high aver age age of about 37 years. They have the highest monthly pay of any cluster, and around 29 percent of their workforce is self-employed, rendering employees the primary group. In this cluster legal professionals, finance professionals, but also stationary plant and machine workers are represented most frequent.

Older, male workers with low educational attainment ***(Cluster 4)***:
This fully male cluster indicates an elderly demographic with an average age of roughly 50. They have finished formal education at a lower level (on average approximately 4). They earn a middling monthly salary on average, and 43 percent of them work for themselves. This cluster characteristically comprises workers in mining, as truck drivers or mechanics.

Young, working men with middle-paying jobs ***(Cluster 5)***:
With an average age of just 29, this cluster is younger and predominately made up of males. They have completed their formal education to a middling level (on average, approximately 5). None of them are self-employed, and their average monthly compensation is in the middle. They mostly work as employees. Similar to this clusters female counterpart, most individuals in this cluster work in sales or elementary jobs, with typical male professions, such as protective services and mining also being represented highly.

RTI and SML per Cluster

Figure 4.13 and Figure 4.14 depict the mean SML- and RTI-values per cluster. Additionally, Table 4.7 presents bivariate regression coefficients of an individual belonging to a particular cluster on the RTI and SML. Three core observations emerge:

	Cluster 5	Cluster 2	Cluster 0	Cluster 1	Cluster 4	Cluster 3
RTI	0.154***	0.281***	−0.567***	−0.142***	0.092***	0.021**
SML	0.056***	−0.179***	0.211***	0.001	−0.122***	0.038***

Table 4.7: Bivariate Regression Coefficients of Cluster on RTI and SML (***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$, No asterix: No significant association)

Observation 1:
Cluster 0 (High-earning male professionals) shows the most extreme values for both SML and RTI, and they are opposite in direction. As the highest educated and highest earning cluster, they have the lowest mean-RTI and the highest value for SML. This is in line with the observations made in the prior subsections, making this cluster the candidate cluster for high SML and low RTI. This suggests that employees within this cluster was and continues to be relatively unaffected by rule-based AI automation, however might see high task replacement in the future, given their low rule-based but high ML-enabled automatibility.

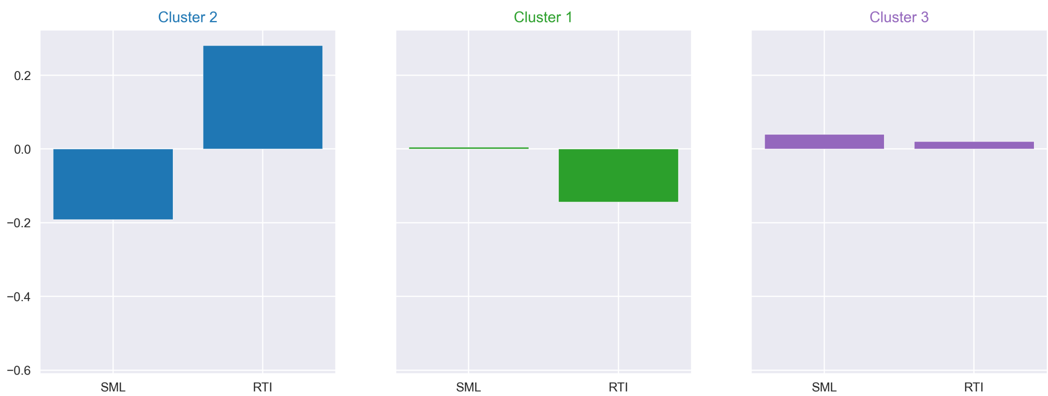


Figure 4.13: Scores Second Three Socio-Demographic Clusters Colombia (source: (World Bank 2014), own calculations)

Observation 2:
Cluster 2 and Cluster 4, which are very similar in profile and mainly distinguish themselves in gender, also show similar tendencies towards automation. The older and less educated workers show above average RTIs and below average SML-scores, with a larger effect size for the female cluster 2. While both clusters have a positive RTI and negative SML, Cluster 2 has both the largest RTI and the smallest SML of all clusters. Being the cluster with the lowest monthly pay, Cluster 2 might be a survivor of prior automation waves, keeping its competitive advantage through cost-efficiency. Occupations prevalent in this cluster might have experienced degrading employment quality over the last few years, however, do not seem too endangered through ML.

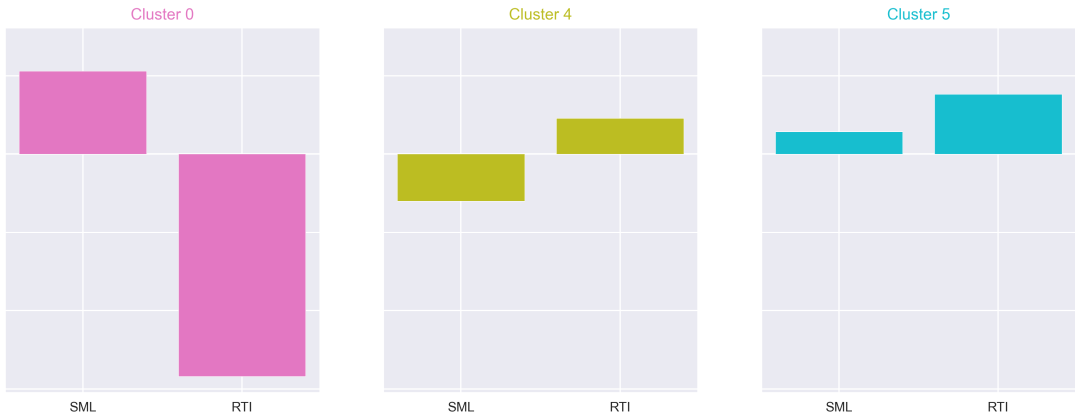


Figure 4.14: Scores Second Three Socio-Demographic Clusters Colombia (source: (World Bank 2014), own calculations)

Observation 3:
Cluster 1, 3, and 5 which taken together depict the average young to middle-aged workforce with a secondary degree and are mostly distinct in gender and share of self-employment all show close to average SML, with little deviation in between the clusters. However, between Cluster 1 and Cluster 5 differences can be observed for RTI, which in the distribution seemed to not differ by gender at first. However, within the younger workforce male workers show higher RTI than female workers, which overlaps with prior findings of Autor and Dorn (2013) and Webb (2020) for developed labour markets, in their case however applied to the middle-aged group.

4.3. Patterns in Other Countries and Interim Summary

This section serves to mention whether patterns deviated or were replicated in our other countries of analysis. For this, first the insights of general socio-demographic analysis as performed for Colombia (to be found in Appendix D) and the country-specific insights of the cluster analyses (to be found in Appendix E) are summarised. Generally it can be noted that prior observations hold water in the analysis over different countries, as has been seen in the global analysis (See: Section 4.1).

4.3.1. Socio-Demographic Analysis

As observed for Columbia, two primary observations for RTI, on being negatively correlated to income and educational attainment, and increasing slightly with age, proved to be robust over our countries of analysis. However, in some countries, contrary to Colombia, women do indeed lower higher, aligning with the insights of Goos, Manning, and Salomons (2014) and Webb (2020). Yet, this observation differs from country to country, and does not meet the level of robustness as for advanced economies. For SML, the patterns of increase with wage and education also were replicated. With only one exception, Vietnam, female workers have consistently higher SMLs than male workers.

4.3.2. Cluster Analysis

For the following countries, certain trends were consistently replicated. Similar to Observation 1 in the prior section, the cluster which showed the highest income or education, or both, typically have a significantly higher SML than RTI. Regardless of country, the observed trend that RTI is for those with lower education while it is low for those with higher education and vice versa for SML holds true. The results of the cluster analysis entrench this insight. However, the "high-performer" clusters did not always prove to have the highest average SML. For certain economies, such as the Yunnan province of China, younger female workers with high education but yet little income are most exposed to Machine Learning and only have little less than average exposure to digitisation. Their male counterparts perform similarly in the ration between SML and RTI, but score considerably more desirable. Generally, for the Yunnan Province, even though all clusters are exposed to automation technologies, the female clusters generally are exposed more to both digitisation and ML. For RTI, this is in contrast to the third Observation 3 and prior findings by Goos, Manning, and Salomons (2014) and Webb (2020), leaving women not only inhibited through existing barriers, but also overproportionately endangered through automation of all kinds.

This pattern of deviation from advanced economies can be observed among the other considered economies too. In Sri Lanka, the first two clusters 1 and 0 (See Appendix E), also have considerably higher exposure to both digitisation and ML than their comparable male clusters 3 and 6. It is relevant to note here that the higher exposure to routine-task automation particularly affects higher aged female workers. Compared to younger female clusters, which also typically show higher levels of educational attainment, the older female clusters without exception demonstrate higher RTI. For men, this is not necessarily the case; only in China and Laos do old male workers have higher RTIs than young male workers, and also only if they work low-paying jobs. Particularly free of vital insight is the analysis of employment status. When analysed over different countries, no particular patterns or reoccuring observations emerged.

4.3.3. Interim Summary

This subsection serves to gather the main insights from the results section, gathered in lower depth. Through this, this subsection also serves as a connector to the discussion section, and as a summary of the answers to our sub-research question.

Despite individual socio-demographic groups and clusters showing considerable differences in their exposure to automation, the first observation of this summary acknowledges no individual groups but a broad field of economies: Oftentimes, SML and RTI characteristically cover opposite extremes of the same trait. RTI often is higher for men than for women, while SML is higher for women. RTI especially is high for those with low levels of education and low wages, whereas SML is affiliated with higher levels of education and high wages. In terms of occupations, SML covers the high skilled occupations whereas RTI covers the low skilled. A very brief answer to RQ1 could thus be: "close to all occupations are affected through automation of at least one digital automation technology". Before individual social groups are to be considered, the macro-observation must be acknowledged: Digital automation technologies, as expected, cover a much broader array of tasks, occupations, and socio-demographic groups. This is mainly due to a broad range of new tasks which can be conducted by AI, which lets automation reach the knowledge workers and renders SML complementary to RTI for emerging economies.

However, the implications this spread of potentially automatable tasks has are ambivalent. Whether or to what extent high SML actually means automation for knowledge workers is discussed in Section 5.2. The accordingly raised subquestion what implications automation of knowledge tasks has on education systems is discussed in Section 5.3.

Moreover, it could be shown that close to generally, women are more exposed to automation. Female workers consistently have higher SMLs than male workers, even when similar clusters are compared. Whether RTI is higher for women is seemingly dependent on the national context; this thesis' yields no definite answer like prior work on RTI in advanced economies does. Yet overall, when taking RTI and SML into account alike, women rather see themselves covered through digital automation than men. This is further discussed in its implications on social policy and gender equality in Section 5.4.

Finally, it could be robustly observed that, for the same levels of education, old workers have slightly higher RTIs than younger workers. For SML, young workers seem to have higher SML in the clusters with lower education. However, the "high performer" cluster with high education and high income, which characteristically were middled aged, also had high SML, which evens this observation out for the higher educated clusters, where higher age normally also went along with higher SML.

Discussion

Whereas the sub-research questions of this thesis have been addressed in Chapter 4, this chapter devotes attention back to the overarching research question, bringing attention to what insights this thesis' brings for social and economic policymakers in emerging economies. The first section of this chapter considers the breadth of affected tasks, the observation that a very large share of tasks could be conducted through robots, computers and AI in the future. The second and third section consider the newly affected knowledge workers, where the second section reviews exposure of knowledge workers with a critical lense, postulating different underlying dynmics than with blue-collar workers, and the third discusses the falling promise of the social ladder behind education. Fourth, automation of women, who in emerging economies see considerably higher exposure than in advanced economies is discussed against the social background of gender equality. The fifth section summarises and puts together a set of policy recommendations, transferring insights from this thesis to what had been discussed in Section 2.3.3. Finally, the sixth section uncovers this thesis limitations.

5.1. RTI, SML and the Breadth of Tasks to be Replaced

A frequently emerging pattern is that for some clusters RTI and SML show differences in polarity, often observed for older workers with lower education and for observed every time for the high-earning high-educated clusters. The results for instance show that for older workers RTI was significantly above-average, rendering them susceptible to automation of routine tasks rather than by AI whereas high-educated workers rather seeing their tasks be AI-automatable. While socio-demographic differences will be considered in higher detail in the following subsections, this subsection focuses on the more high-level observation: Through affecting tasks and workers not affected by routine automation, the introduction of AI will increase the share of tasks that can be automated, taking in new tasks, mainly knowledge or skill-based non-routine tasks. This overlaps with the findings of Frey and Osborne (2013) just like the hypotheses in Brynjolfsson and Mitchell (2017) and subsequent findings by Brynjolfsson, Mitchell, and Rock (2018).

While this does not suffice to predict detailed developments, it justifies theoretical concern for emerging economies: As discussed in the mechanism of leapfrogging in Section 2.3.2 and the underlying emerging threat, especially as both automation of routine tasks and AI-enabled automation are still to be expected for many emerging economies with the potential to happen simultaneous, automation can take place more broadly and extensively, magnifying potential consequences. In Section 2.3.3 the comparatively low capability of effective policy design in emerging economies was discussed; put briefly, emerging economies are disadvantaged through more complex and sudden diffusion processes of technology with more complex stakeholder structures which threatens exploitational developments and increases uncertainty. Uncertainty further is magnified through informality. Simultaneously, educational and vocation infrastructure is limited and thus prone to underperform in the design of supply-side policies, when compared to advanced economies. With the establishment of a widely-available digital infrastructure, routine-automating and AI technologies would see a simultaneously opened window for entry into individual emerging economies, further intensifying the effects of technological progress. The finding that this double-diffusion of automation technologies affects a broader share of workers and tasks only worsens expectations on the magnitude of rises in unemployment and share of people affected by wage-suppression.

However, a broad diffusion of automation technologies, although connotated with justified concern, will have nuanced effects and should not be understood as fully absent of benefit. Some concerns typical for emerging economies could potentially be effectively adressed through automation in general and especially AI-enabled automation. Two problems, braindrain, the outflux of the most skilled workers, and gaps in labour markets which cannot be filled through low availability of particular skills could be beneficially adressed if not solved. Moreover, if exploitational practices from foreign actors from advanced economies can be tackled effectively, gains in productivity can cause increased welfare, economic competitiveness and even push development, reducing the gap to advanced economies.

5.2. Automation Reaching Knowledge Workers, but to Which Effect?

Our results highlight that also for emerging economies automation potentials change significantly in the advent of AI, affecting new tasks, occupations and segments of the labour market. The most robust observation is that the prior unaffected groups, those with high income and education, are suddenly more prone to be automated. This not only aligns with the tone of prior warnings that the share of affected occupations will grow largely with AI, spilling over to occupations which require high education (Frey and Osborne 2013), but also with the most recent insights gained on generative AI for the US labour market (Brynjolfsson, Li, and Raymond 2023, Eloundou et al. 2023). When considering LLMs such as GPT-4, high-education tasks are expected to be replaced over two pathways: as per our analysis, because an AI can take their tasks over, or because secondly, tasks can be guided by LLMs, reducing the required experience, or in some cases required education (Brynjolfsson, Li, and Raymond 2023).

This high potential for automation of tasks in a novel part of the labour market however is not to be directly interpreted with a reduction in labour demand for high-education and high-wage tasks. In fact, while the classical observation for automation through robots and computers is reduction of employment share and wage depression (Autor, Levy, and Murnane 2003, Goos, Manning, and Salomons 2009, Webb 2020 and many others), more recent observations have shown that high AI exposure of knowledge workers is associated to wage increases, specifically for employees (Fossen, Samaan, and Sorgner 2022). An explanation for this is that highly educated workers rather profit off productivity increases than employees with low education. After all, workers with academic degrees in white collar occupations often enjoy higher independence and more project-based than task-based work, rendering them capable to use upcoming AI tools to their own benefit. One benefit could be increasing their own contribution, expanding the goals of their work, whereas for blue-collar workers or white-collar workers with more fixed tasks, productivity increases through automation oftentimes mean that less of their time and wage is required to reach their goal, rendering them more replacable.

Other knowledge workers affected, but not threatened by automation might be speicalised workers in high demand, or even shortage, and some bottlenecks in their task composition, such as medical doctors (7th highest SML) and nurses (3rd highest SML) (Frey and Osborne 2013, Webb 2020). High demand and lack of automatability in some tasks could keep these occupations employed even if all automatable tasks were to be replaced, because their tasks are specialised enough to not be taken over by other professions in the same institutions and redistribution of non-automated tasks would take place on few individuals. The decision to reduce employment for such occupations would require either very high progress in automation, lack of financial means, or the acceptance to let their employers underperform, which in the special case of hospitals would be hardly justifiable.

However, not all high-wage and high-education professions are protected from loss in employment share through their nature of occupation. Particularly those that require a high level of specialisation but have a sizable number of tasks that can be automated. This includes some technical, financial, and legal professions where AI and machine learning models like GPT-4 can draft documents, perform in-depth analyses, or even forecast outcomes based on massive amounts of data. These occupations are not just characteristic for our most robust and SML cluster, the high-performers, but also threatened by market effects. The tasks of those engineers, lawyers or stock brokers which work on a more task-based approach, promise high financial gain through automation, especially through system which can be replicated close to indefinitely (Frey and Osborne 2013, Ernst, Merola, and Samaan 2019). Yet, the delay of automation induced by institutional factors remains uncertain. In the US, these occupations - while highly susceptible - are continuing to grow as of June 2023 (McKinsey Company 2023), which implies that also in emerging economies there will be no significant automation affecting knowledge workers in the short-term future.

5.3. Inequality and the Social Ladder of Education

This section also considers high-paying educations, typically those of ”knowledge workers”. However when compared to the preceding section, this sections’ emphasis is put less on the likelihood of ”disappearance” of these occupations, but more on how they will change as a means of social mobility and how wealth equality and wealth accumulation will likely be affected.

The share of national income allocated to labour is declining consistently in advanced as in emerging economies (). Out of many explanations, technological progress is one of the most popular (This comes as no surprise given that this represents core neoclassical argumentation). The underlying dynamic is that as the productivity of capital increases (through technological progress allowing the conduction of tasks priorly performed by humans now at a lower price) labour loses productivity in relation to capital: Using one unit of labour promises less gain than the use of one unit of capital, so capital prevails over labour and the share of labour in total income declines (For more detail see: 2). (Hicks 1932)

With respect to social justice, or equality of opportunity, this dynamic implies that with progress in technology which substitutes human labour, it becomes more and more difficult for citizens to participate in the economy through employment, as long as no new forms of employment are generated (Piketty and Zucman 2014). This thesis’ analysis of course is not capable of producing insight on the emergence of new forms of employment, but casts light on another aspect of this dynamic, particular for emerging economies: The differences in how occupations just as social groups score in RTI and SML suggest that with the avent of AI, a broader field of tasks will indeed be substituted, taking away some of the available tasks from workers, aligned with the projection for advanced economies by Brynjolfsson and McAfee (2014). In fact, not only do the results suggest that more occupations will be affected, but also that different social levels will be affected. High-wage occupations which can be entered through high educational attainment, a cornerstone behind the promise of social mobility, might still be reachable, but are likely to yield less advancement on the social ladder, and thus are less of a mean to reach financial independence.

Admittedly, also in advanced economies younger generations are hardly considered to have a realistic opportunity to become wealthy through labour (Piketty and Zucman 2014). For many young people and in low and middle-wage occupations, inhibitors to reaching a financially comfortable positions are the decreases of affordability of general needs, like housing, health services or since 2022 food (European Commission 2022, International Labour Organization 2022, Parliament 2022). If the occupations which as per status quo were the "means-to-go" for social mobility will increase in substitutability with capital too, upwards mobility were to be further constrained (or exclusive to those born with capital). On the other side, little logic suggests that wealth accumulation would be constrained for owners of capital, rather the opposite. In a "market-let-loose" scenario, following the logic of free market economics, there is no mechanism to ensure a "fair" distribution of income and consequently wealth (Brynjolfsson and McAfee 2014), and with pathways closing for young people, little reasoning beyond personal development remains for higher educational aspirations. The implication for distribution of wealth would not be novel – rather continuing following the "success to the successful"- archetype – but would through the introduction further burdens on accumulation of wealth for those without capital, suggest a more and more polarised distribution process; not between high- and low-income, but between the majority of population and few owners of capital.

For emerging economies, which also already see a decline in the labour share of income (Dao, Das, and Koczan 2020), there is little difference to the core dynamic except for the theoretical balancing mechanism of trade, the Stolper-Samuelsen Theorem, which benefits workers in emerging economies through their competitive advantage in price, providing employment through offshoring (). However, this countering effect is yet to proven empirically, more recent studies show a decline in labour share with little balance through "imported" employment. (Arora and Gambardella 2004, Bhagwati, Panagariya, and Srinivasan 2004, Bhagwati, Panagariya, and Srinivasan 2004, Dao, Das, and Koczan 2020). Effects that remain are the potential future creation of new and demanded tasks and occupations in favour of maintaing the labour share, but also weaker labour market institutions than compared to advanced economies (with the exception of the United States)

5.4. The Future of Work for Women - Hurdles to Equality?

Our global results align with prior findings of Autor, Levy, and Murnane (2003), Goos, Manning, and Salomons (2014) and Webb (2020) that in terms of RTI, and hence the susceptibility of robotic or computer-based automation, women generally show lower values. They also align with prior findings that these conditions change when considering newer, more sophisticated AI models, to a state in which women slow slightly higher automation potentials (Webb 2020, Georgieff and Hyee 2021, McKinsey Company 2023). However, in the comparison over countries it became apparent that the beneficial effect for women to have lower automation potential for routine tasks, is not as robust as was demonstrated by researchers for labour markets of advanced economies, whereas the higher SML remained stable.

For gender equality in labour markets, this might impose serious (even if not definite) implications. Especially as women in emerging economies often are subject to additional barriers in progression in and especially entry into the workforce (International Labour Organisation 2018) the shift in automation potentials with upcoming AI diffusion could exacerbate existing inequalities.

At first, attention should be devoted to the lack in robustness of RTI over countries, meaning that digitalisation, or robotisation, might not necessarily mean the automation of characteristically male professions on a global scale, but only for advanced economies (Autor, Levy, and Murnane 2003, Goos, Manning, and Salomons 2014 and Webb 2020) and few emerging economies. While an equality induced by economisation of male jobs may not the desirable path, a spread of inequality in countries where women not only are disadvantaged in wage and access to labour markets but also through cuts in their labour demand seems even worse. Regarding the projection of an even worse inequality, it is relevant to additionally take into account the findings of Maloney and Molina (2016), Messina (2016) and Das and Hilgenstock (2018), which suggest that automation for routine jobs has neither yet taken place in the same extent in emerging as it has for advanced economies nor that it takes place in the polarising pattern as for advanced economies. Instead, according to the RTI results in this thesis we would rather expect digitalisation sweeping jobs of the low-educated, which would render the implications on inequality darker-looking than for advanced economies, as those affected by digitalisation and robotisation tend to be more economically vulnerable, and face higher entries into other occupations. Women in these more vulnerable groups, who often partake in precarious employment (Young 2010) and face higher burdens of unpaid care work in emerging than in advanced economies (Bustelo, Flabbi, and Viollaz 2019, International Labour Organization 2023), would find themselves most endangered by economic precarity.

For routine service occupations, which often are more female (International Labour Organisation 2018), and already have seen automation in advanced economies (Acemoglu and Restrepo 2022), employment shares could drop drastically with increasing availability of digital infrastructure. Not only would this reduce economic welfare through a drop in labour force participation caused by a conventionally desired step in economic development, but would prevent female independence through employment, and thus income. In fact, for low and lower-middle income countries the gaps in employment and labour force participation – the jobs gap – plays a more significant role to the difference in labour income between genders than the gender pay gap does (International Labour Organization 2023). Interestingly, women leaving the labour force without actively searching for new jobs, because their old occupation was economised and entry barriers into new occupations might be either too high or their entry might not be supported, would neither appear in statistics for wage disparities or unemployment. If this would occur, the reduction in equality would thus remain a lot less visible than phenomenons like the gender pay gap, which however does not mitigate any harm on individual womens and their households finances. In summary, the lack of robustness for RTI to be higher in males implies a threat of additional precarity to women in those countries, where RTI is higher for women.

Secondly, women are likely to be disadvantaged compared to their male peers because of their characteristically higher SMLs across clusters of all age groups. They might on one hand face higher barriers into and more competition by AI in higher-level positions and on the other see more competition through higher influxes into the more social professions (Frey and Osborne 2013, UNESCO 2023), with high uncertainty regarding future changes in demand. Regarding higher-level positions, development of AI might widen the already existing representational gap – women are less represented in high-level positions and aspire less for them due to social expectations and norms – even more, entrenching gender inequality. Beyond the higher average SML three dynamics, some taken from feminist literature, are believed to affect the future of women in leadership positions with respect to automation: First, the tasks related to many higher-level positions, also characteristically with high SMLs, might be less demanded as ML-technologies become more common. Second, the opportunities for training and professional development may also be biased. There is a risk that skill adaptation training may be disproportionately offered to those already in leadership roles who are largely male (Chin et al. 2008) as companies invest in training their workforce to adapt to the new AI-driven environment . Third, this new AI-related difficulty may exacerbate the long-standing obstacles women face in advancing in the corporate world, including biases in promotions, the difficulty juggling work and family obligations, and a lack of mentorship or networking opportunities (De Beauvoir 1949, Chin et al. 2008). If the tasks that women are more likely to perform in these positions become automated, it might result in women being either forced out of these positions or being overseen for promotions because their skill set is viewed as ”redundant” by the new technological standards (UNESCO 2023).

5.5. Policy Considerations

As digital infrastructure is made available in an economy-wide extent, labour markets in emerging economies bear the potential to be more suddenly and broadly affected by automation than advanced economies' labour markets. The social implications of both AI-enabled automation and digitalisation are profound, affecting not just the skills required to participate in a rapidly evolving job market but also a broad range of socio-demographic groups. On the other hand, the backward advantage – if leveraged effectively – promises highly desired ”catch-up” potentials. Accordingly, policymakers in emerging economies must proactively prepare policy measures to prepare and equip their labour forces for impending transitions. In sections 2.2.3 and 2.3.3, policy areas for advanced and emerging economies in particular were identified to prepare labour forces for automation.

The first policy area, described as ”supply-side” policies by Autor, Mindell, and Reynolds (2022), consists of policy levers aiming to change the supply of labour; the skills of workers. Limited by comparatively weaker vocational and educational systems, supply-side policies are best enforced creatively, exceeding the limited means provided by established institutions. Special limitations arise when workers who barely can stem leaving paid employment are expected to devote large amounts of time, which they otherwise would spend acquiring income. Simultaneously, increasing availability of digital infrastructure grants potentials to increase educational efficiency. This thesis proposes a combination of equipping and boosting workers with means made available by digital infrastructure (Olejniczak, Śliwowski, and Leeuw 2020), preceded by government funded programmes to increase digital literacy. Equipping and boosting means to first provide information on what skills will be relevant in the future and to accordingly provide access to digital training platforms, such as Coursera, sidelined by secondly ensuring that those offers are known about and made available to as many as possible.

Vocational and educational systems require mid- to long-term transformation, directed towards fostering the skills of the future (in this thesis' analysis, one of the least SML skills was to interact with humans). Digital literacy, which was not accounted for in the STEP-survey and thus out of our methodological scope, is yet recommended to also be of high emphasis in the future education systems. Moreover, there is a need for programs within educational and vocational systems that inform about the potential impact of automation on their employment. Such programs aid workers in anticipating changes in their respective fields, enabling them to make more informed career choices.

The second policy area discussed by Autor, Mindell, and Reynolds (2022) addresses the ”demand-side” of labour and the restructuring of income distribution. As mentioned in Section 2.3.3, this is particularly difficult for emerging economies, as data necessary to understand how new technologies would decrease the labour share of income is scarce, especially in sectors with high informality. Moreover, this thesis' insights with relevance to demand-side policies, mainly the strengthening of capital reducing opportunities for high paying employment (Section 5.3), do not allow for particularly nuanced policy design. Yet, the strengthening of capital abilities through the broadness of automation maintains its importance for tax policy (Autor, Mindell, and Reynolds 2022), as reductions in labour tax income can negatively affect national balance sheets while reduced binding to workers increases capital mobility; the capability of firms to change locations in case of capital gains tax increases (Zucman 2014, Saez and Zucman 2019). This imposes a large challenge for the governments of economies which are not in a leading position of business attractivity, like the US. While addressing this issue to a satisfactory extent would exceed labour or social policy and thus the scope of this thesis, one particular policy avenue which simultaneously might ameliorate supply-side and demand-side circumstances is worthy of consideration: Programs to better path the way for workers affected by job losses or entry barriers into entrepreneurship. Workers in this position can be better equipped to either start their own enterprises or work as freelancers who provide services by leveraging digital tools and platforms. This could encourage innovation and economic diversity in addition to offering a substitute for traditional employment. Yet, it requires far more thorough research to grasp the potential of demand-side policies in emerging economies.

R&D or innovation system oriented policies is the third and last policy area Autor, Mindell, and Reynolds (2022) discussed for emerging economies. For emerging economies, the AI R&D policy area is considered the least fruitful, as the discrepance in technical development of digital capital and AI between the advanced and emerging economies is too large to be compensated for (See Section 2.3.3). R&D is yet not to be entirely disregarded; rather to be directed wisely. While globally competitive AI systems, such as ChatGPT, might not be the most effective focus, development of smaller AI systems aiming to ease national problems or research on the application of AI systems in the national context promises higher returns of welfare. Regarding applications, scholars of labour economics and business administration alike could divert research towards more sector- or business-specific analysis of how an optimal allocation of available digital capital and labour resources could look like, and what AI systems might be of highest societal and economic value.

This redirection of R&D in fact bridges the gap to the augmentation of policy measures for emerging economies (See Section 2.3.3), in which crucial considerations by Dutz, Almeida, and G. (2018) were brought to the emerging economy context. Given the potentials to incite economic growth in emerging economies promised by digital infrastructure investments, the question of technical diffusion is widely understood to not be a question of *whether* but of *when* and *how*. The *when* and *how* can be influenced through technical diffusion policies (Dutz, Almeida, and G. 2018). Government-funded and, importantly, equally distributed high-speed internet rollout programs, paired with digital literacy programs aid a more smooth transition into an inclusive digital economy. Moreover, inclusivity could be fostered through government-funded data center and cloud computing server programs, made available and taught to to smaller and medium sized companies who normally would experience disadvantage through market entry barriers. Technical diffusion policies as such can serve as the cornerstone for our first and second policy area describe above, but also for our last policy area, product market policies aimed to enable competitive environments, ensuring a more even distribution of newly gained economic welfare (Dutz, Almeida, and G. 2018).

In summary, as the benefits and risks of digital transformation are on the horizon for emerging economies, it remains uncertain whether the broad negative social effects can be cushioned, with no net positive outcomes be guaranteed regardless of adopted measures. Any effort towards a desirable outcome requires a well-balanced implementation of policy steering adoption into the free market while cushioning a sudden wave of automation. A strategy as such is subject to high uncertainty, especially against the broad potentials of digital automation, affecting the majority of occupations and socio-demographic clusters. Yet, the path in which digital technologies diffuse into markets bears an important lever space to preemptively design the landscape of their effects. For emerging economies, especially those where the economy-wide roll-out of digital infrastructure is in its infancy, a comprehensive policy approach allows to steer the advent and diffusion of digital automation onto more desirable paths. (Only) If the diffusion of digital technologies is directed effectively and sidelined by measures to foster digital literacy, entrepreneurial know-how, and competition, economic growth could be stimulated while maintaining or even improving employment and wage equality, promoting local firms and innovation increasing an economies capability to address further issues.

5.6. Limitations

This thesis deserves special attentions in its limitations, of which there are many. The limitations encompass theoretical assumptions in the field of labour economics and this thesis in particular, over this thesis' methodology, the data available for this thesis, interpretations made and the vast uncertainty surrounding the topic of this thesis, rendering the derivation of direct and tangible policy implications very difficult.

1. The principle of capital labour substitution which justifies the main concern behind automation is simplified, and does not account for bounded rationality of actors, decision preferences of employers, is built on neoclassical assumptions which many economists do not agree with. This puts the foundation of not just this thesis but also many other economists works into question.

2. The field of automation in labour economics is dominated by researchers researching from and on advanced economies. While this scientific gap provided the motivation for this thesis, it also caused lack of theoretical background, empirical observations and data availability for emerging economies, rendering research on emerging economies significantly more difficult and limited.

3. The Data taken for the calculation of task indices describes the work of individuals and occupations on too superficial level too obtain a comprehensive image of all tasks conducted. Furthermore, the data provides no indication of how important tasks weigh in their descriptive potential towards the work of individuals or occupations. This does not just affect the validity of this thesis' results, but also of all other who used the STEP-survey or PIAAC to calculate automation indices and task measures.

4. Misinterpeting the research procedure in a parallel research project with the ILO caused me to commit a grave mistake in my time management delaying the calculation for two months. This mismanagement in return imposed strong time constraints in writing and putting together this document. In my personal opinion this document does not reflect the time awarded to the thesis project, and I believe I could have done much better.

5. Another limitation arises from the choice of cross-country-comparison and the choice of countries (emerging economies) itself. Considering multiple countries dictates the non-consideration over factors which are especially specific to individual countries. In this thesis' case the most relevant ignored factor were the differences in labour protection regulation, which shape employer capabilities to substitute workers for technology dramatically. But also the choice of emerging economies, despite being well-motivated, imposed constraints on this analysis, conducted by an author who has never set foot in these countries. A sound feeling of what the level of development in the analysed countries actually is on what economic implications this has is outright absent.

6. What also, through data, time and knowledge constraints, was left out of scope are cultural differences and how they shape the labour markets of their respective countries. The analysed Yunnan province for instance belongs to communist China. The assumption that workers are replaced through AI through the same market mechanisms as would take place in other more capitalist countries is a clear oversimplification.

7. A large, though necessary, simplification is the scoping out of institutional factors which prevents actual in-depth modelling of automation in labour markets. It is unclear whether such a complex environment could ever be modeled and simulated in a satisfying extent, the research filed of labour economics yet has to bring up an extensive model of institutional factors. However, such models would be a necessary step to draw conclusions of predictive nature, and the modelling process of institutional factors – if performed diligently and correctly – would be a promising opportunity to gain understandings over more disaggregate dynamics and behaviours.

Conclusion

This thesis addresses the largely uncharted territory of emerging economies' susceptibility to automation. A literature analysis was conducted with special focus on research on the US and OECD labour markets. An examination of technology life cycles and the differences between technological diffusion in advanced and emerging economies supplements the analysis how susceptible occupations and socio-demographic groups are to automation in emerging economies. As the first primary contribution, Routine Task Intensity, which measures the effects of automation through computers and robotics, and Suitability for Machine Learning (SML), which evaluates the prospective impact of AI, were calculated for emerging economies. This contribution, providing indicators for routine and AI automation potentials comparable across multiple emerging economies which partook in the STEP survey is the first and only one of its kind. Subsequently RTI and SML were analysed across tasks, occupations, and socio-demographic groups, casting light on who is susceptible to automation of what kind. A sociodemographic cluster analysis further endeepened this picture in the second contribution: highlighting patterns of susceptibility across diverse socio-demographic segments and their social implications.

This thesis finds that, beyond the consequences of routine task automation, the introduction of AI will dramatically increase the range of jobs, vocations, and socio-demographic categories susceptible to automation. Taken together with leaps in technological diffusion, and the circumstance that the necessary infrastructure for both digitalisation and AI still needs to be made available in an economy-wide fashion in many emerging economies, this justifies the expectation of not just particularly broad but also sudden waves of automation in emerging economies. This suddenness and broadness per nature can be understood as exacerbators to the already high uncertainty of policymaking aimed to smoothen societal effects of automation. This thesis further finds that both low- and high-educated workers show increased susceptibilities to automation, although by different technical means. This raises important issues about how educational systems —traditionally seen as engines for social mobility— are changing in their function. Furthermore, there were clear gender differences in automation susceptibility, with women, particularly those in service roles, being more vulnerable to AI-induced automation. Men in developing markets were not always overproportionately vulnerable to robot-led automation, in contrast to observations in industrialised economies. While this does not exacerbate inequality per se, it does deprive women of a competitive advantage on the path to equality.

Further research and governmental attention is necessary to deepen the understanding of automation in emerging. The procedure of calculating the task indicators RTI and SML has unveiled relevant limitations in the usage and interpretation of such indicators for the data available on emerging economies. The creation of a robust model for the institutional dynamics of automation, potentially created in cooperation with scholars of business administration, and the compilation of detailed data of occupational task composition in emerging economies are needed. The results of this thesis underline the criticality of a proactive, educated, and, most importantly, comprehensive policy set in order to anticipate threats and take advantage of the opportunities that automation presents in emerging economies.

References

Acemoglu, D. and D. Autor (2011). “Skills, tasks and technologies: Implications for employment and earnings”. en. In: *Handbook of labor economics*. Vol. 4. Elsevier, pp. 1043–1171.

Acemoglu, D., A. X. He, and D. Le Maire (Mar. 2022). *Eclipse of Rent-Sharing: The Effects of Managers’ Business Education on Wages and the Labor Share in the US and Denmark*. Working Paper No. 29874. National Bureau of Economic Research. URL: `https://www.nber.org/papers/w29874`.

Acemoglu, D. and P. Restrepo (2018). “The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment”. en. In: *American Economic Review* 108.6, pp. 1488–1542.

Acemoglu, D. and P. Restrepo (2022). “Tasks, automation, and the rise in us wage inequality”. In: *Econometrica* 90.5, pp. 1973–2016.

Anwar, M.A. and M. Graham (2021). “Between a rock and a hard place: Freedom, flexibility, precarity and vulnerability in the gig economy in Africa”. en. In: *Competition Change* 25.2, pp. 237–258.

Arntz, M., T. Gregory, and U. Zierahn (2016). “The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis”. en. In: *OECD Social, Employment and Migration Working Papers* 189.

Arntz, M., T. Gregory, and U. Zierahn (2017). “Revisiting the Risk of Automation”. en. In: *Economics Letters* 159, pp. 157–160.

Arora, A. and A. Gambardella (2004). “The globalization of the software industry: Perspectives and opportunities for developed and developing countries”. In: *Innovation Policy and the Economy, Volume 5*. MIT Press, pp. 1–32.

Atkinson, A. (2008). “De Benedetti Lecture Series”. en. In.

Autor, D. (2015). “Why are there still so many jobs? The history and future of workplace automation”. en. In: *Journal of economic perspectives* 29.3, pp. 3–30.

Autor, D. (2022). “An Inclusive Future? Technology, New Dynamics, and Policy Challenges”. In: *Brookings Institution*. URL: `https://www.brookings.edu/wp-content/uploads/2022/05/Inclusive-future_Technology-new-dynamics-policy-challenges.pdf`.

Autor, D. and D. Dorn (2009). “This job is” getting old”: Measuring changes in job opportunities using occupational age structure”. en. In: *American Economic Review* 99.2, pp. 45–51.

Autor, D. and D. Dorn (2013). “The growth of low-skill service jobs and the polarization of the US labor market”. en. In: *American economic review* 103.5, pp. 1553–97.

Autor, D., D. Dorn, and G. Hanson (2013). “The China syndrome: Local labor market effects of import competition in the United States”. In: *American economic review* 103.6, pp. 2121–2168.

Autor, D. and M. Handel (2013). “Putting tasks to the test: Human capital, job tasks and wages”. In: *Journal of Labor Economics* 31.S1, S59–S96.

Autor, D., L. Katz, and M. Kearney (2008). “Trends in U.S. Wage Inequality: Re-Assessing the Revisionists”. en. In: *Review of Economics and Statistics* 90.2, pp. 300–323.

Autor, D., L. Katz, and A. Krueger (1998). “Computing Inequality: Have Computers Changed the Labor Market?” en. In: *Quarterly Journal of Economics* 113I, pp. 1169–1214.

Autor, D., F. Levy, and R. Murnane (2003). “The Skill Content of Recent Technological Change: An Empirical Exploration”. en. In: *Quarterly Journal of Economics* 118.4, pp. 1279–333.

Autor, D., D. Mindell, and E. Reynolds (2022). *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. Cambridge: MIT Press.

Bhagwati, J., A. Panagariya, and T. Srinivasan (2004). “The Muddles over Outsourcing”. In: *Journal of Economic Perspectives* 18.4, pp. 93–114.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. Springer. ISBN: 978-0-387-31073-2.

Black, S. and A. Spitz-Oener (2010). “Explaining Women’s Success: Technological Change and the Skill Content of Women’s Work”. In: *The Review of Economics and Statistics* 92.1, pp. 187–194.

Bresnahan, T. and M. Trajtenberg (1995). “General purpose technologies ‘Engines of growth’?” In: *Journal of econometrics* 65.1, pp. 83–108.

Brynjolfsson, E., D. Li, and L. Raymond (2023). *Generative AI at work*. Tech. rep. National Bureau of Economic Research.

Brynjolfsson, E. and A. McAfee (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. en. WW Norton Company.

Brynjolfsson, E. and T. Mitchell (2017). “What can machine learning do? Workforce implications”. en. In: *Science* 358.6370, pp. 1530–1534.

Brynjolfsson, E., T. Mitchell, and D. Rock (May 2018). “What can machines learn, and what does it mean for occupations and the economy?” en. In: 108, pp. 43–47.

Burris, B. (1993). *Technocracy at work*. SUNY Press.

Burris, B. (1998). “Computerization of the workplace”. In: *Annual review of Sociology* 24.1, pp. 141–157.

Bustelo, M., L. Flabbi, and M. Viollaz (2019). *The Gender Labor Market Gap in the Digital Economy*. Tech. rep. Inter-American Development Bank. URL: https://publications.iadb.org/en/gender-labor-market-gap-digital-economy.

Carbonero, F., J. Davies, et al. (2023). “The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam”. In.

Carbonero, F., E. Ernst, and E. Weber (2020). “Robots worldwide: The impact of automation on employment and trade”. In: *IAB Discussion Paper*. URL: https://doku.iab.de/discussionpapers/2020/Dp072020.pdf.

Card, D. and T. Lemieux (2001). “Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis”. en. In: *Quarterly Journal of Economics* 116, pp. 705–746.

Carneiro, P. and Sokbae L. (2009). ““Trends in Quality-Adjusted Skill Premia in the United States, 1960-2000”. en. In: *CEMMAP Working Paper* P02.09.

Chin, J. et al. (2008). *Women and leadership: Transforming visions and diverse voices*. John Wiley & Sons.

Creswell, J.W. (2020). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (Fifth*. en. Pearson.

Crowley, M. et al. (2010). “Neo-Taylorism at work: Occupational change in the post-Fordist era”. In: *Social Problems* 57.3, pp. 421–447.

D’ignazio, C. and L. F. Klein (2020). *Data feminism*. MIT press.

Dao, M. C., M. Das, and Z. Koczan (2020). “Why is labour receiving a smaller share of global income?” In: *Oxford Academic*. URL: https://academic.oup.com/economicpolicy/article-abstract.

Das, M. and B. Hilgenstock (2018). *The Exposure to Routinization: Labor Market Implications for Developed and Developing Economies*. Tech. rep. International Monetary Fund. URL: https://www.imf.org/en/Publications/WP/Issues/2018/06/13/The-Exposure-to-Routinization-Labor-Market-Implications-for-Developed-and-Developing-45930.

Davis, S. (1992). *“Cross-Country Patterns of Changes in Relative Wages” NBER Macroeconomic Annual, 86*. en. Cambridge: MIT Press, pp. 239–292.

De Beauvoir, S. (1949). “The second sex”. In: *Social Theory Re-Wired*. Routledge, pp. 346–354.

Dobbe, R., T. K. Gilbert, and Y. Mintz (2021). “Hard Choices in Artificial Intelligence”. In: *Artificial Intelligence* 300 (C), p. 103555. DOI: https://doi.org/10.1016/j.artint.2021.103555.

Du, Y. and A. Park (2018). *Changing Demand for Tasks and Skills in China*. Background Report. For the World Bank Group (WBG)-Development Research Center under the State Council (DRC) Report on New Drivers of Growth in China. HKUST Institute for Emerging Market Studies, Hong Kong.

Dutz, M. A., R. K. Almeida, and Packard T. G. (2018). *The Jobs of Tomorrow: Technology, Productivity, and Prosperity in Latin America and the Caribbean*. World Bank. DOI: 10.1596/978-1-4648-1222-4. URL: https://elibrary.worldbank.org/doi/epdf/10.1596/978-1-4648-1222-4.

Eloundou, T. et al. (2023). "Gpts are gpts: An early look at the labor market impact potential of large language models". In: *arXiv preprint arXiv:2303.10130*.

Ernst, E., R. Merola, and D. Samaan (2019). "Economics of artificial intelligence: Implications for the future of work". In: *IZA Journal of Labor Policy* 9.1.

European Commission (2022). "Inflation increases poverty unevenly, widening gaps across the EU". In: URL: https://joint-research-centre.ec.europa.eu/inflation-increases-poverty-unevenly-widening-gaps-across-the-EU.

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

Ferry, W.H. (1964). *The Triple Revolution*. en. Santa Barbara: Center for the Study of Democratic Institutions.

Firpo, S., N. M. Fortin, and T. Lemieux (2011). *Occupational Tasks and Changes in the Wage Structure*. Tech. rep. IZA Discussion Paper.

Fitzenberger, B. and K. Kohn (June 2006). *Skill Wage Premia, Employment, and Cohort Effects: Are Workers in Germany All of the Same Type?* en. University of Freiburg Working Paper.

Fossen, F., D. Samaan, and A. Sorgner (2022). "How Are Patented AI, Software and Robot Technologies Related to Wage Changes in the United States?" In: *Frontiers in Artificial Intelligence* 5, p. 869282.

Fossen, F. and A. Sorgner (2019). "Mapping the future of occupations: transformative and destructive effects of new digital technologies on jobs". en. In: *Форсайт* 13.2 (eng, pp. 10–18.

Frey, C.B. (2019). "The technology trap". en. In: *The Technology Trap*. Princeton University Press.

Frey, C.B. and M.A. Osborne (2013). "The future of employment: How susceptible are jobs to computerisation?" en. In: *Technological fo recasting and social change* 114, pp. 254–280.

Georgieff, A. and R. Hyee (2021). "Artificial intelligence and employment: New cross-country evidence". In.

Gerschenkron, A. (1962). "Economic backwardness in historical perspective". In: *Cambridge MA*.

Goos, M. and A. Manning (2007). "Lousy and lovely jobs: The rising polarization of work in Britain". In: *The Review of Economics and Statistics* 89.1, pp. 118–133.

Goos, M., A. Manning, and A. Salomons (2009). “Job Polarization in Europe”. In: *American Economic Review* 99.2, pp. 58–63.

Goos, M., A. Manning, and A. Salomons (2014). “Explaining Job Polarization: Routine-Biased Technological Change and Offshoring”. In: *American Economic Review* 104.8, pp. 2509–2526.

Green, F. (2012). “Employee Involvement, Technology and Evolution in Job Skills: A Task-Based Analysis”. In: *Industrial and Labor Relations Review* 65.1, pp. 36–67.

Hardy, W., R. Keister, and P. Lewandowski (2018). “Educational upgrading, structural change and the task composition of jobs in Europe”. In: *Economics of Transition* 26.2, pp. 201–231.

Hémous, D. and M. Olsen (2022). “The rise of the machines: Automation, horizontal innovation, and income inequality”. In: *American Economic Journal: Macroeconomics* 14.1, pp. 179–223.

Hicks, J. (1932). *The theory of wages*. Springer.

Homsombath, K. (2020). *How COVID19 could accelerate digitalization in Laos, United Nations Development Programme*. en. URL: https://www.la.undp.org/content/lao_pdr/en/home/blog/2020/how-covid19-could-acceleratedigitalization-in-laos.html..

International Finance Corporation (2020). *Artifical Intelligence in Emerging Markets: Opportunities, Trends, and Emerging Business Models*. en. URL: fc.org/wps/wcm/connect/95a40480-27b5-4b99-8b4c-7768ae6a53a2/AI-Report_Web.pdf?MOD=AJPERES&CVID=nhLrsRc.

International Labour Organisation (2018). “International Women’s Day: ILO: Women still less likely to be active in the labour market than men in most of the world”. In: *International Labour Organization*. Accessed: [Your Access Date]. URL: https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_619119/lang--en/index.htm.

International Labour Organization (2008). *International Standard Classification of Occupations: ISCO-08*. Accessed: 2023-09-01. URL: https://isco-ilo.netlify.app/en/isco-08/.

International Labour Organization (2022). “Global Employment Trends for Youth 2022”. In: URL: https://www.ilo.org/global-employment-trends-youth-2022.

International Labour Organization (2023). *New data shine light on gender gaps in the labour market*. Tech. rep. Accessed: [15/08/2023]. International Labour Organization. URL: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_870519.pdf.

Iyer, A. (2018). “Moving from Industry 2.0 to Industry 4.0: A case study from India on leapfrogging in smart manufacturing”. In: *Procedia Manufacturing* 21, pp. 663–670.

Juma, C. (2016). *Innovation and its enemies: Why people resist new technologies*. Oxford University Press.

Katz, L., G.W. Loveman, and D. Blanchflower (1995). *“A Comparison of Changes in the Structure of Wages in Four OECD Countries*. en. Ed. by Richard Freeman and Lawrence Katz. Differences and Changes in Wage Structures National Bureau of Economic Research. University of Chicago Press.

Katz, L. and K.M. Murphy (1992). “Changes in relative wages, 1963–1987: supply and demand factors”. en. In: *The quarterly journal of economics* 107.1, pp. 35–78.

Keynes, J.M. (1933). *Economic possibilities for our grandchildren*. en.

Kochhar, R. (2015). “America’s Shrinking Middle Class: A Close Look at Changes Within Metropolitan Areas”. In: URL: `https://www.pewresearch.org/social-trends/2016/05/11/americas-shrinking-middle-class-a-close-look-at-changes-within-metropolitan-areas/`.

Kristal, T. (2013). “The capitalist machine: Computerization, workers’ power, and the decline in labor’s share within US industries”. In: *American Sociological Review* 78.3, pp. 361–389.

Lane, M. and A. Saint-Martin (2021). “The Impact of Artificial Intelligence on the Labour Market: What Do We Know So Far?” en. In: *OECD Social, Employment and Migration Working Papers* 256. DOI: `10.1787/7c895724-en`. URL: `https://dx.doi.org/10.1787/7c895724-en`.

Lehner, J. (2017). *Job Polarization, 2016 Update — oregoneconomicanalysis.com*. `https://oregoneconomicanalysis.com/2017/03/31/job-polarization-2016-update/`. [Accessed 10-Jul-2023].

Leondes, Cornelius T (2001). *Expert systems: the technology of knowledge management and decision making for the 21st century*. Elsevier.

Levy, F. (2018). “Computers and populism: artificial intelligence, jobs, and politics in the near term”. In: *Oxford Review of Economic Policy* 34.3, pp. 393–417.

Lewandowski, P. et al. (2022). “Technology, skills, and globalization: Explaining international differences in routine and nonroutine work using survey data”. en. In: *The World Bank Economic Review* 36.3, pp. 687–708.

Luxburg, U. (2007). “A tutorial on spectral clustering”. In: *Statistics and computing* 17.4, pp. 395–416.

MacQueen, J. (1967). “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14, pp. 281–297.

Maloney, W. F. and C. A. Molina (2016). *Are Automation and Trade Polarizing Developing Country Labor Markets?* Tech. rep. World Bank. URL: `https://documents1.worldbank.org/curated/en/579161509027951402/pdf/WPS7922.pdf`.

McKinsey Company (2023). “Generative AI and the future of work in America”. In: *McKinsey & Company*. Accessed: [15/08/2023]. URL: `https://www.mckinsey.com/mgi/our-research/generati`.

Messina, J. (2016). “Routinization and Employment: Evidence for Latin America”. In: URL: `https://www.cedlas.econo.unlp.edu.ar/wp-content/uploads/doc_cedlas276.pdf`.

Michaels, G., A. Natraj, and J. Van Reenen (2014). “Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years”. In: *Review of Economics and Statistics* 96.1, pp. 60–77.

Murphy, K.M., W.G. Riddell, and P.M. Romer (1998). “Wages, Skills and Technology in the United States and Canada”. en. In: *General Purpose Technologies*. Ed. by E. Helpman. Cambridge, MA: MIT Press.

Murtagh, F. and P. Legendre (2014). “Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion?” In: *Journal of classification* 31.3, pp. 274–295.

Olejniczak, K., P. Śliwowski, and F. Leeuw (2020). “Comparing behavioral assumptions of policy tools: Framework for policy designers”. In: *Journal of Comparative Policy Analysis: Research and Practice* 22.6, pp. 498–520.

Parliament, Hansard - UK (2022). “Cost of Living: Public Well-being”. In: URL: `https://hansard.parliament.uk/Lords/2022-10-20/debates/cost-of-living-public-well-being`.

Perez, C. (2003). *Technological revolutions and financial capital*. Edward Elgar Publishing.

Piketty, T. and G. Zucman (2014). “Capital is Back: Wealth-Income Ratios in Rich Countries 1700-2010”. In: *Quarterly Journal of Economics* 129.3, pp. 1255–1310. DOI: `10.1093/qje/qju018`.

Polanyi, M. (1966). “The logic of tacit inference”. en. In: *Philosophy* 41.155, pp. 1–18.

Porter, Michael E et al. (1998). *Clusters and the new economics of competition*. Vol. 76. 6. Harvard Business Review Boston.

Precht, R.D. (2022). *Freiheit für Alle: Das Ende der Arbeit wie wir sie kannten*. de. Goldmann Verlag.

R., Douglas (2009). “Gaussian Mixture Models”. In: *MIT Lincoln Laboratory*. URL: `http://leap.ee.iisc.ac.in/sriram/teaching/MLSP_16/refs/GMM_Tutorial_Reynolds.pdf`.

Reimers, N. and I. Gurevych (2019). “Sentence-BERT: Sentence embeddings using siamese BERT-networks”. In: *arXiv preprint arXiv:1908.10084*.

Saez, E. and G. Zucman (2019). *The triumph of injustice: How the rich dodge taxes and how to make them pay*. WW Norton & Company.

Skott, P. and F. Guy (2005). “Power-biased technological change and the rise in earnings inequality”. In: *Journal of Economic Inequality* 3.3, pp. 227–254.

Skott, P. and F. Guy (2007). “A model of power-biased technological change”. In: *Economics Letters* 92.1, pp. 124–131.

Spitz-Oener, A. (2006). “Technical change, job tasks, and rising educational demands: Looking outside the wage structure”. In: *Journal of Labor Economics* 24.2, pp. 235–270.

Stefano, V. (2015). “The rise of the just-in-time workforce: On-demand work, crowdwork, and labor protection in the gig-economy”. en. In: *Comp. Lab. L. Pol'y J* 37, p. 471.

Tinbergen, J. (1975). *Income Distribution: Analysis and Policies*. en. Amsterdam: North-Holland Publishing Company.

UNESCO (2023). “The effects of AI on the working lives of women”. In: URL: https://www.unesco.org/articles/effects-ai-working-lives-women.

Vallas, S. P. (1993). *Power in the workplace: The politics of production at AT&T*. SUNY Press.

Waterman, D. A. (1985). *A guide to expert systems*. Addison-Wesley Longman Publishing Co., Inc.

Webb, M. (2020). *The Impact of Artificial Intelligence on the Labor Market*. en. Working Paper,

World Bank (2014). *STEP skills measurement surveys: innovative tools for assessing skills*. fr.

Young, M.C. (2010). “Gender Differences in Precarious Work Settings”. In: URL: https://www.researchgate.net/publication/227627744_Gender_Differences_in_Precarious_Work_Settings.

Zucman, G. (2014). “Taxing across borders: Tracking personal wealth and corporate profits”. In: *Journal of economic perspectives* 28.4, pp. 121–148.

Appendix A: ISCO-08

Code	Occupation	Skill Level
1	Managers	3, 4
2	Professionals	4
3	Technicians and Associate Professionals	3
4	Clerical Support Workers	2
5	Service and Sales Workers	2
6	Skilled Agricultural, Forestry, and Fishery Workers	2
7	Craft and Related Trades Workers	2
8	Plant and Machine Operators, and Assemblers	2
9	Elementary Occupations	1
0	Armed Forces Occupations	1, 2, 4

Table A.1: ISCO 1-Digit Codes with Corresponding Occupations and Skill Levels (source: International Labour Organization 2008

Term	Definition
Job	A set of tasks and duties performed, or meant to be performed, by one person, including for an employer or in self-employment.
Occupation	The kind of work performed in a job. It is a set of jobs whose main tasks and duties are characterized by a high degree of similarity. A person may be associated with an occupation through the main job currently held, a second job, a future job, or a job previously held.
Skill	The ability to carry out the tasks and duties of a job.

Table A.2: Definitions as per ISCO-08 guidelines (source: International Labour Organization 2008)

Appendix B: Method: Routine Task Intensity

STEP Survey Item	Composite Task Measure	Count
Tasks that involve 30 or more minutes of thinking?	Non-Routine Cognitive	10021
Does your work require the use of Statistical or other type of analysis?	Non-Routine Cognitive	10019
Does your work require the use of CAD software?	Non-Routine Cognitive	10019
Does your work require the use of designing websites?	Non-Routine Cognitive	10019
Does your work require the use of presentation or graphics software?	Non-Routine Cognitive	10019
Does your work require the use of accounting or financial software?	Non-Routine Cognitive	10019
Does your work require the use of macros and complex equations?	Non-Routine Cognitive	10019
Does your job require the use of spreadsheets?	Non-Routine Cognitive	10020
Does your job require word processing?	Non-Routine Cognitive	10020
Does your job require searching info on the internet?	Non-Routine Cognitive	10020
Does your job require using email?	Non-Routine Cognitive	10020
Freedom to decide in which way you work	Non-Routine Cognitive	10021
Does your work require the use of software programming?	Non-Routine Cognitive	10019
Operate or work with heavy machines?	Non-Routine Cognitive	10021
Does your work require the use of managing computer networks?	Non-Routine Cognitive	10019
Drive a car, truck or three wheeler?	Non-Routine Cognitive	10021
Does your job require reading instruction manuals?	Non-Routine Cognitive	10030
Advanced math at work?	Non-Routine Cognitive	10030

STEP Survey Item	Composite Task Measure	Count
Does your job require to read reports?	Non-Routine Cognitive	10030
Repair electronic equipment?	Non-Routine Cognitive	10021
Does your job require reading newspapers, magazines, or books?	Non-Routine Cognitive	10030
Have you ever written anything else at work?	Non-Routine Cognitive	10030
Time involved with customer	Non-Routine Interpersonal	10020
Direct and check the work of other workers	Non-Routine Interpersonal	10021
Contact with people non-coworkers?	Non-Routine Interpersonal	10021
Do you regularly use a phone?	Non-Routine Interpersonal	10021
Operate or work with heavy machines?	Non-Routine Manual	10021
Drive a car, truck or three wheeler?	Non-Routine Manual	10021
Repair electronic equipment?	Non-Routine Manual	10021
Does your job require handling bills?	Routine Cognitive	10030
Perform any other multiplication or division at work?	Routine Cognitive	10030
Use or calculate fractions or decimals at work?	Routine Cognitive	10030
Does your job require data entry?	Routine Cognitive	10020
Measure sizes, weights, distances at work?	Routine Cognitive	10030
Does your job require filling out forms?	Routine Cognitive	10030
Lifted anything that weighed more than 50 pounds at work?	Routine Manual	10021
Do you regularly use a bar code reader?	Routine Manual	10013

Table B.1: STEP Survey Items and Composite Task Measures for extensive RTI calculation

Appendix C: Method: SML

Step Survey Item	SML
Does your job require you to use a bar code reader?	3.74323
Does your job require you to use or calculate fractions, decimals, or percentages?	3.73274
Do you measure or estimate sizes, weights, distances?	3.68577
Do you do data entry?	3.68111
Do you use advanced functions in spreadsheets such as macros?	3.65669
Do you direct and check the work of other workers, or supervise?	3.6554
Do you read very short notes or instructions that are only a few sentences long?	3.65153
Do you write very short notes, lists, or instructions that are only a few sentences long?	3.64476
Do you perform any other multiplication or division?	3.64111
Do you contact customers, clients, students, or the public?	3.62731
Do you read reports?	3.62133
Do you manage computer networks?	3.60176
Do you calculate prices or costs?	3.58101
Do you use spreadsheets?	3.56549
Do you read forms?	3.56227
Do you use email?	3.55345
Do you use a telephone, mobile phone, pager, or other communication device?	3.54979
Does your job require you to use more advanced math, such as algebra, geometry, or trigonometry?	3.54498
Do you search for information on the internet?	3.52859
Do you use databases?	3.52275
Do you make formal presentations to clients or colleagues to provide information or persuade them of your point of view?	3.51856
Do you design websites?	3.51208
Do you read instructions?	3.50675

Step Survey Item	SML
Does your job require you to use software packages for tasks such as designing websites, programming, managing networks?	3.50252
Do you carry out short, repetitive tasks?	3.5024
Do you fill out bills or forms?	3.49983
Do you lift or pull anything heavy?	3.48828
Do you read newspapers, magazines, or books?	3.48105
Do you use presentation or graphics software, such as PowerPoint?	3.46818
Do you drive a car, truck, or three-wheeler?	3.4605
Do you do book-keeping, accounting, or use financial software?	3.43983
Do you learn new things?	3.43748
Does your job require you to use a computer?	3.42402
Do you do statistical analysis or other types of analysis?	3.41835
Do you do software programming?	3.41525
Do you meet or interact with customers, clients, students, or the public?	3.40186
Does your job require you to engage in thinking tasks that take longer than half an hour?	3.39599
Do you have the freedom to decide how to do your work in your own way, rather than following a fixed procedure or a supervisor's instructions?	3.38451
Do you use CAD software (Computer-Aided Design)?	3.29445
Do you read bills or financial statements?	3.2871
Does your job require you to operate or work with any heavy machines or industrial equipment?	3.27553
Do you use word processing software?	3.24581
Do you repair or maintain electronic equipment?	3.22218
Is your job physically demanding?	3.17966

Table C.1: STEP Survey Items and Corresponding SML Values

Appendix D: Results: SML and RTI

D.1. Yunnan (China)

D.1.1. Routine Task Intensity

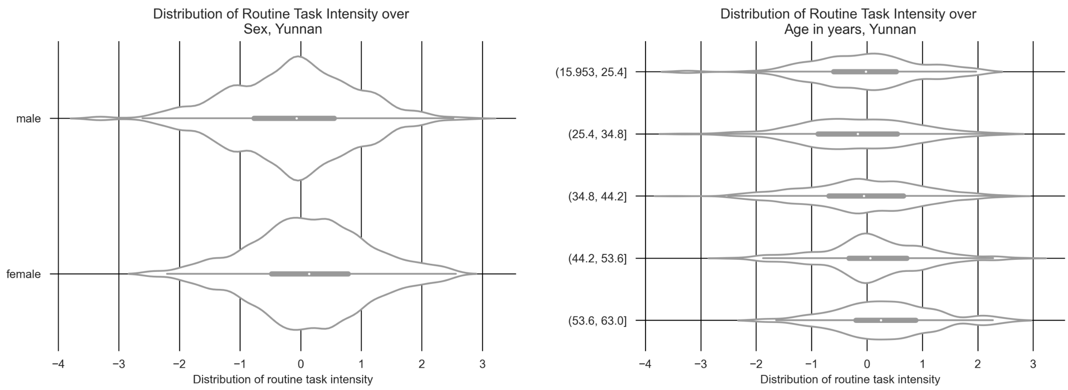


Figure D.1: Relationship between Sex and Age to SML in Yunnan (China)

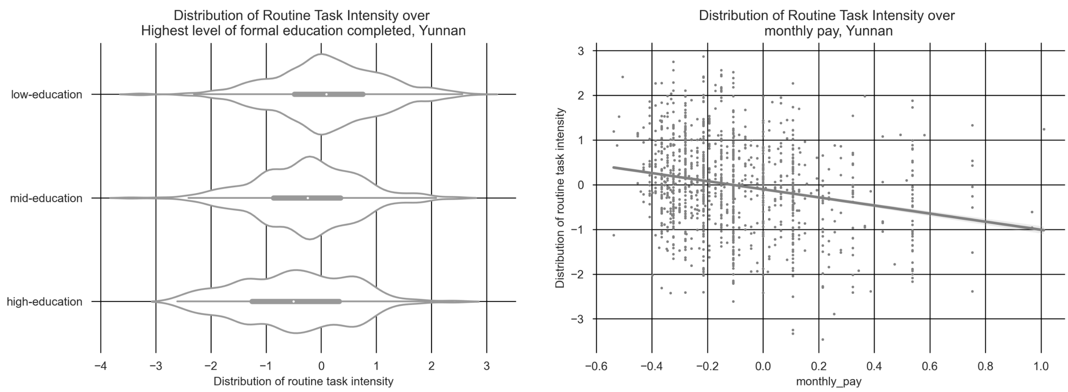


Figure D.2: Relationship between Educational Attainment and Wage to RTI in Yunnan (China)

D.1.2. Suitability for Machine Learning

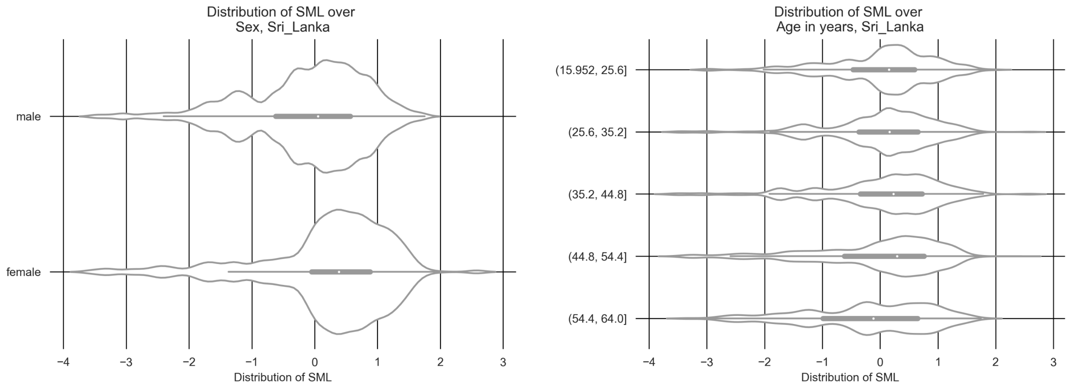


Figure D.3: Relationship between Sex and Age to SML in Yunnan (China)

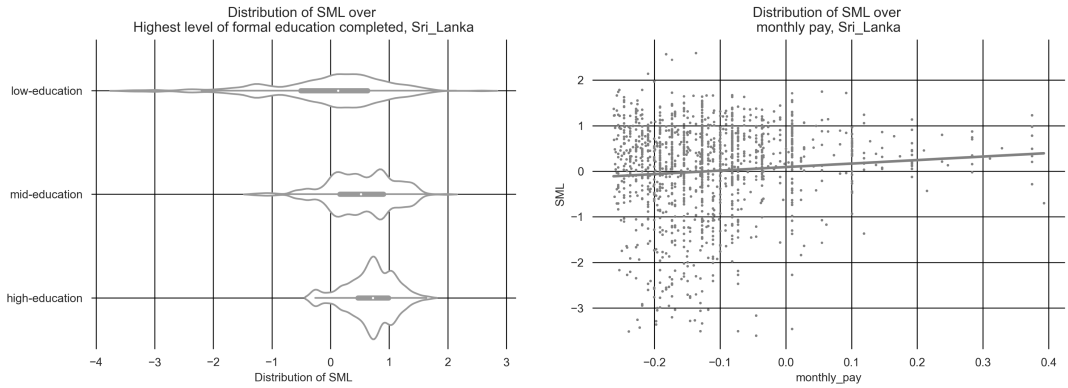


Figure D.4: Relationship between Educational Attainment and Wage to SML in Yunnan (China)

D.2. Vietnam

D.2.1. Routine Task Intensity

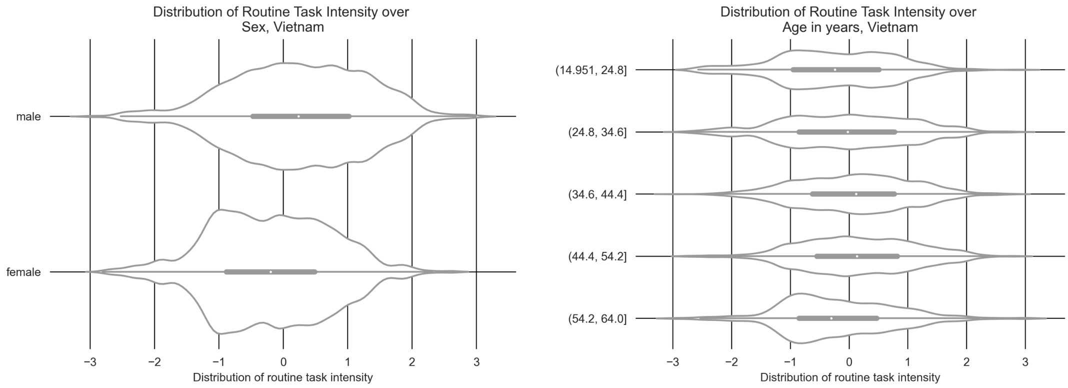


Figure D.5: Relationship between Sex and Age to SML in Vietnam

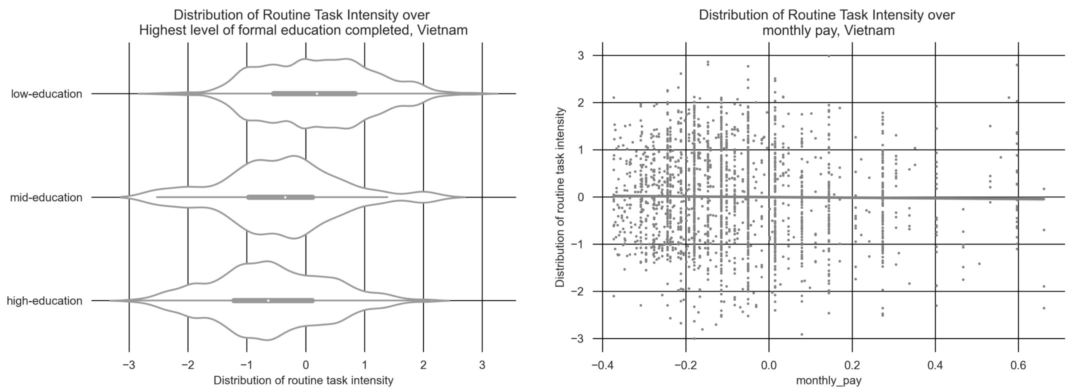


Figure D.6: Relationship between Educational Attainment and Wage to RTI in Vietnam

D.2.2. Suitability for Machine Learning

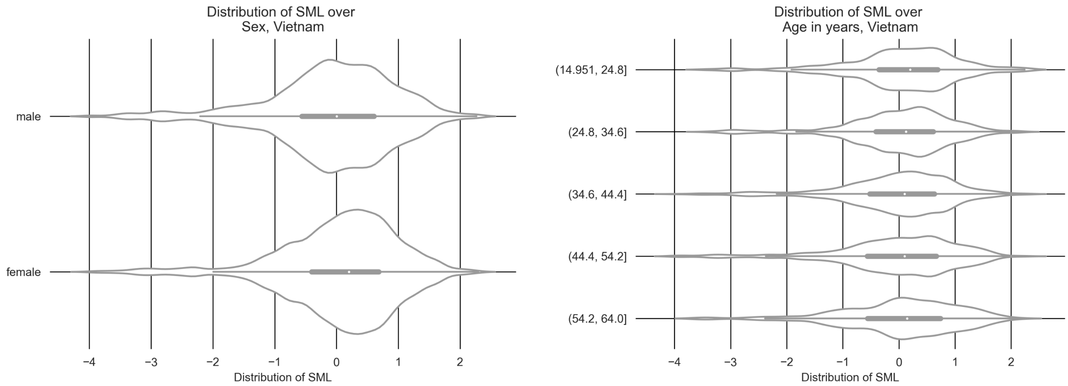


Figure D.7: Relationship between Sex and Age to SML in Vietnam

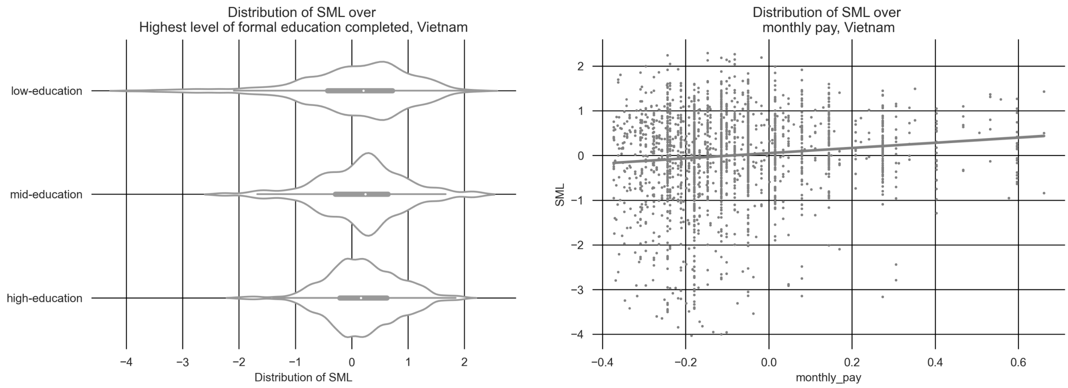


Figure D.8: Relationship between Educational Attainment and Wage to SML in Vietnam

D.3. Laos

D.3.1. Routine Task Intensity

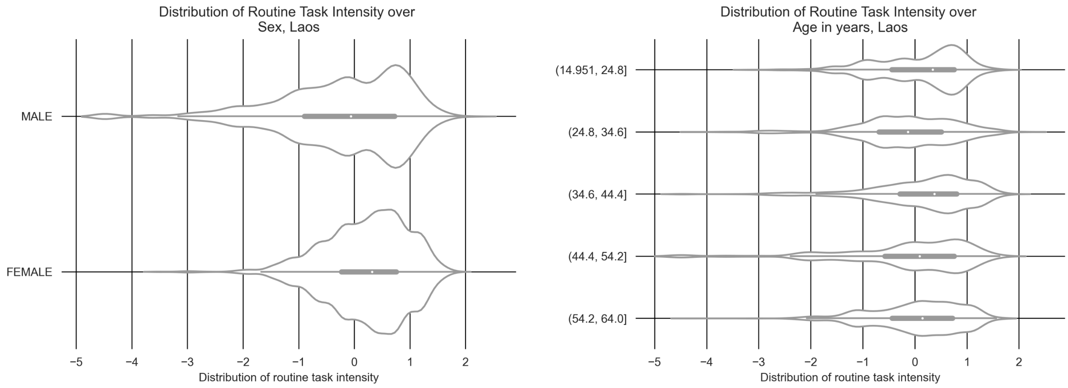


Figure D.9: Relationship between Sex and Age to SML in Laos

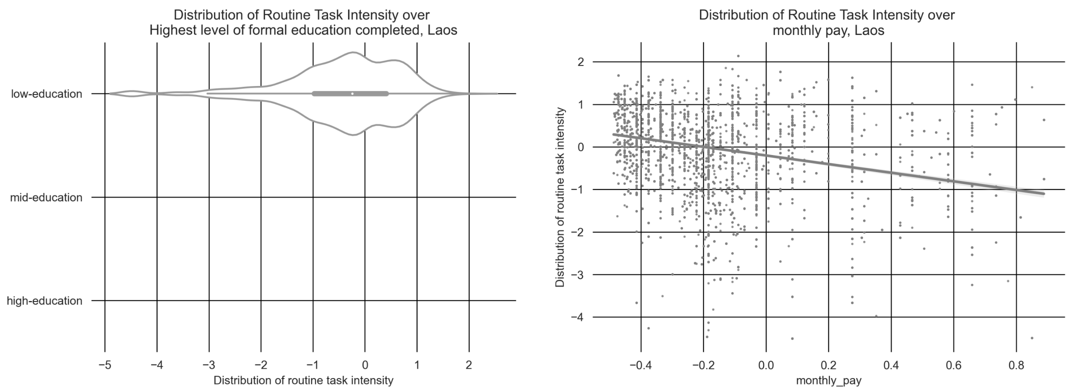


Figure D.10: Relationship between Educational Attainment and Wage to RTI in Laos

D.3.2. Suitability for Machine Learning

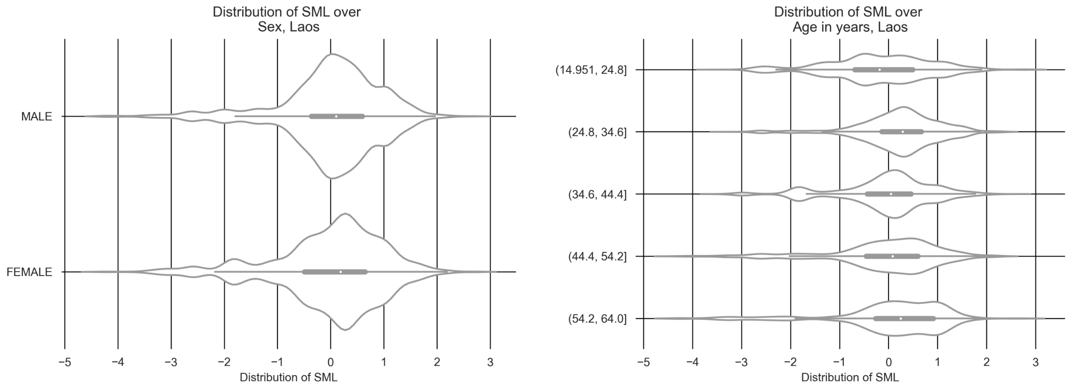


Figure D.11: Relationship between Sex and Age to SML in Laos

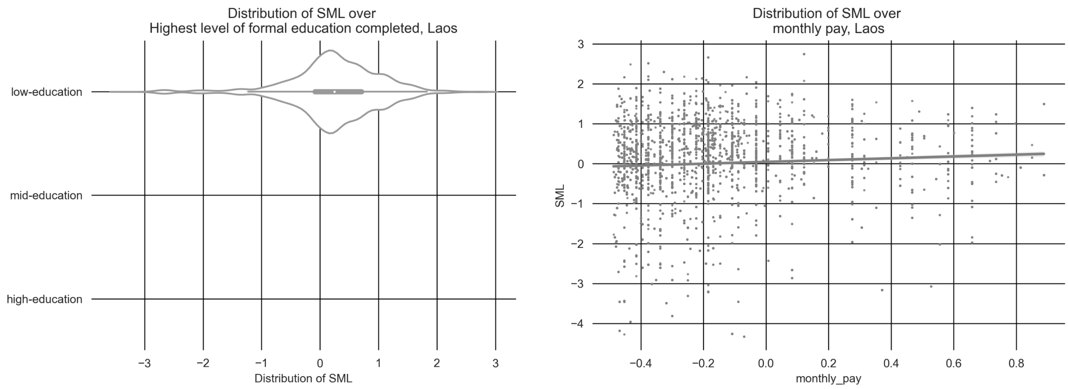


Figure D.12: Relationship between Educational Attainment and Wage to SML in Laos

D.4. Sri Lanka

D.4.1. Routine Task Intensity

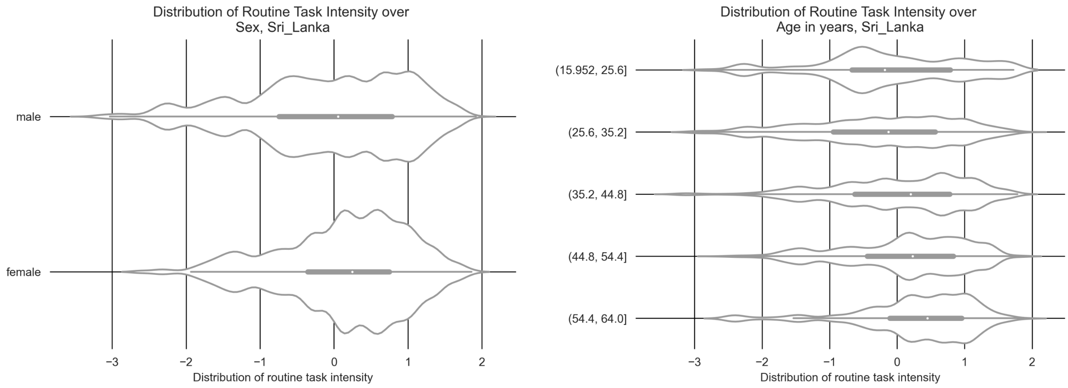


Figure D.13: Relationship between Sex and Age to SML in Sri Lanka

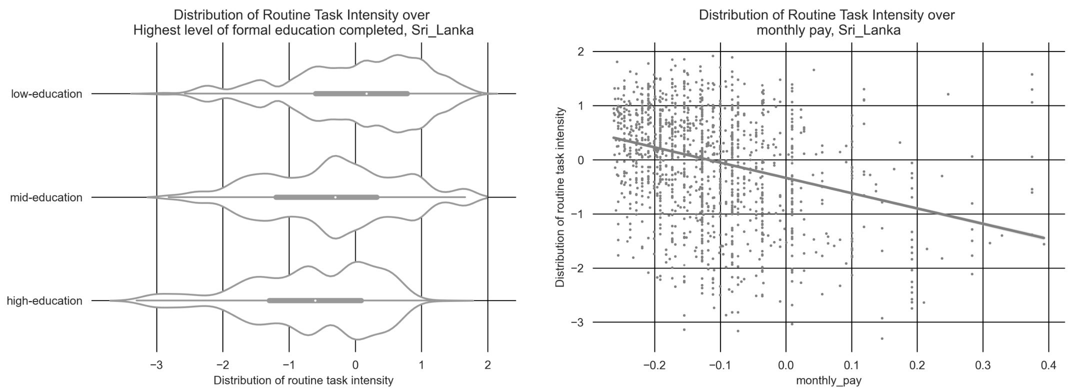


Figure D.14: Relationship between Educational Attainment and Wage to RTI in Sri Lanka

D.4.2. Suitability for Machine Learning

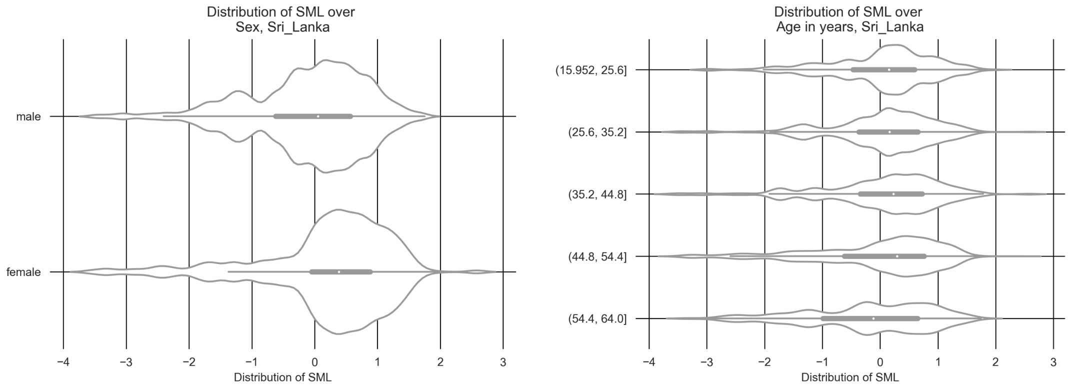


Figure D.15: Relationship between Sex and Age to SML in Sri Lanka

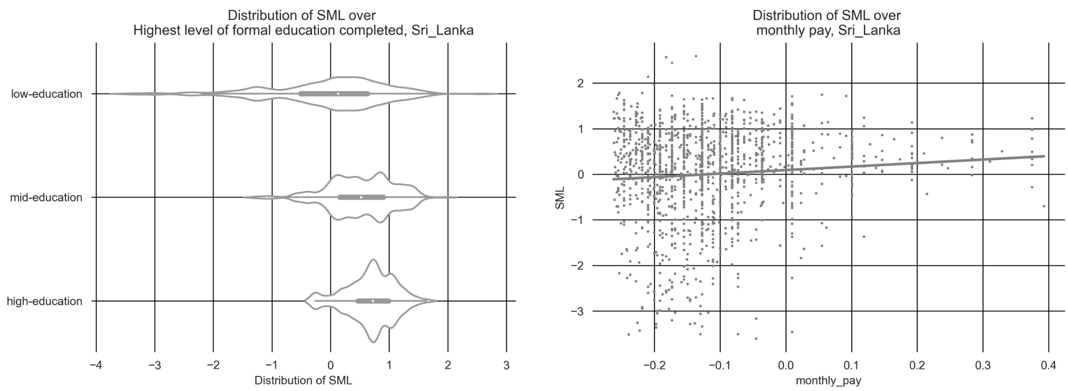


Figure D.16: Relationship between Educational Attainment and Wage to SML in Sri Lanka

Appendix E: Results: Clustering

E.1. Sri Lanka

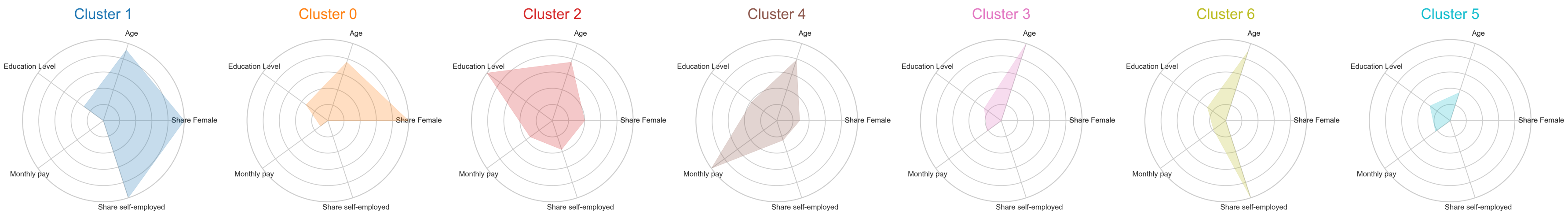


Figure E.1: Profiling Spiderwebs for Clusters in Sri Lanka

Figure E.2: Automation Scores for Cluster in Sri Lanka

E.2. Yunnan (China)

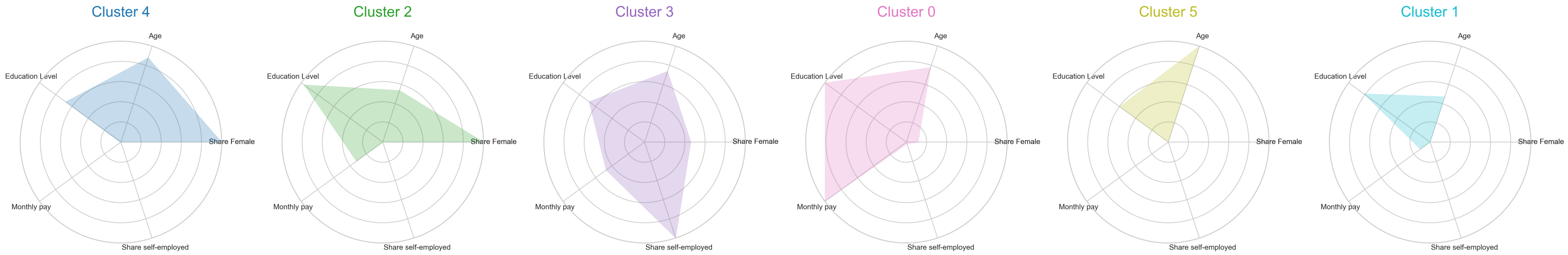


Figure E.3: Profiling Spiderwebs for Clusters in Yunnan (China)

Figure E.4: Automation Scores for Cluster in Yunnan (China)

E.3. Laos

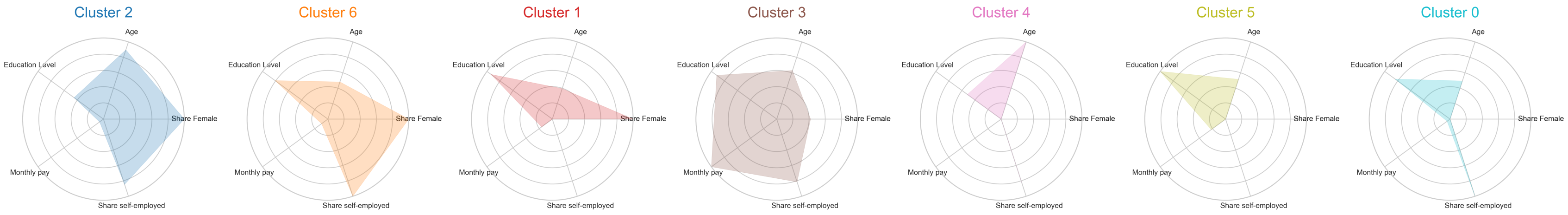


Figure E.5: Profiling Spiderwebs for Clusters in Laos

Figure E.6: Automation Scores for Cluster in Laos

E.4. Vietnam

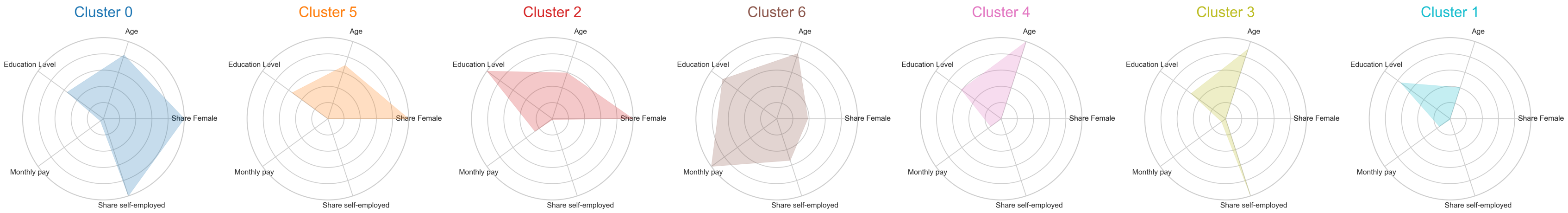


Figure E.7: Profiling Spiderwebs for Clusters in Vietnam

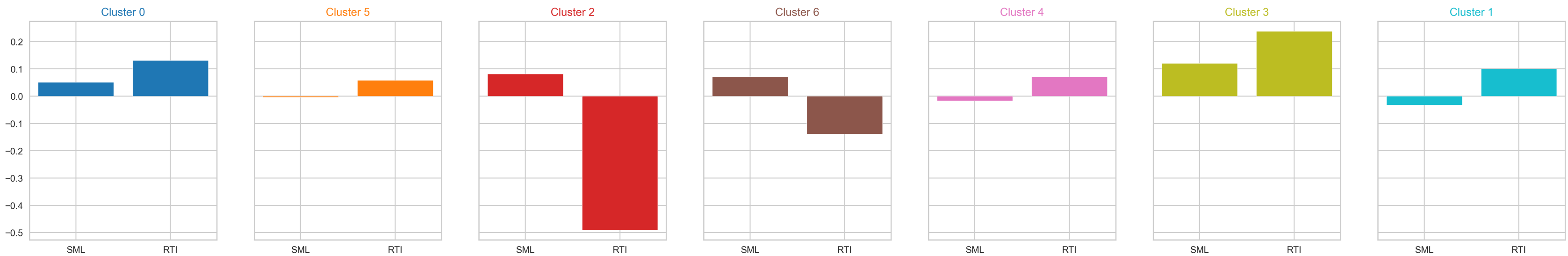


Figure E.8: Automation Scores for Cluster in Vietnam