DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

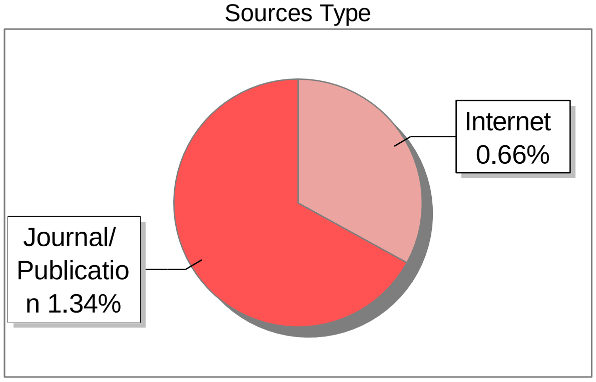
Author Name	CHANDANA B C
Title	RISING UNEMPLOYMENT RATES DUE TO AUTOMATION, ECONOMIC SHIFTS, AND CRISES.
Paper/Submission ID	3577990
Submitted by	charankumarba@gmail.com
Submission Date	2025-05-05 16:05:28
Total Pages, Total Words	8, 3488
Document type	Research Paper

Result Information

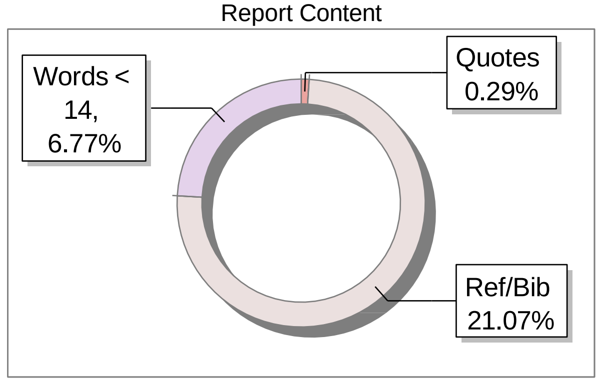
Similarity **2 %**

1 10 20 30 40 50 60 70 80 90

Sources Type



Report Content



Exclude Information

Quotes	Excluded
References/Bibliography	Excluded
Source: Excluded < 14 Words	Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File

DrillBit Similarity Report

2	3	A	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	moam.info	1	Internet Data
2	www.diva-portal.org	1	Publication
3	A Systematic Review of the Effectiveness of Self-Compassion-Related Intervention by Kl-2020	1	Publication

RISING UNEMPLOYMENT RATES DUE TO AUTOMATION, ECONOMIC SHIFTS, AND CRISES.

1st Chandana B C
Department of Information Science
The Oxford College of Engineering
Bangalore, India
chandanaise2022@gmail.com

2nd H C Pranjali Holla
Department of Information Science
The Oxford College of Engineering
Bangalore, India
pranjaliise2026@gmail.com

***Abstract*— This research considers how automation impacts job employment in Southern Africa, where the region was already faced with high unemployment levels. Drawing upon data between the years 2004 to 2017 across ten countries, it discovers that higher automation actually increases unemployment. Yet foreign investment and industrial productivity can contribute towards mitigating losses in employment. The causation is two-sided—automation causes unemployment, and high unemployment further decelerates automation. The research demands policies of education and vocational training to facilitate automation leads to advancement and not exacerbates unemployment.**

***Keywords*—Automation, Artificial Intelligence(AI), Technological Unemployment, Structural Change, Job Displacement, Labor Market Transformation, Upskilling and Reskilling, Fourth Industrial Revolution.**

I. INTRODUCTION

Automation is changing the way we work, thanks to advances in artificial intelligence and robotics. While these technologies can make industries more efficient and even create new types of jobs, they also bring real concerns—especially in places like Southern Africa, where unemployment is already high. Many workers may find their roles replaced by machines, and without the right skills, they could struggle to find new opportunities. The real challenge isn't just how people and economies adjust. By investing in education and training, countries can help their workforce stay relevant and make automation a tool for progress, not exclusion.

Automation is no longer a distant concept—it's a growing force that's already reshaping jobs, industries, and entire economies. With advances in robotics and artificial intelligence, many routine tasks once done by people are now handled by machines. This shift is especially important to understand in regions like Southern Africa, where unemployment is already high and economic structures are less resilient. Ignoring automation's impact could widen inequalities, leaving many behind. At the same time, automation offers opportunities—new jobs, improved productivity, and economic growth—if societies can adapt quickly enough. The topic matters because it sits at the crossroads of technology, employment, and social development. Preparing for these changes is crucial. If managed well, automation can be a powerful driver of progress. But if left unchecked, it could deepen existing challenges. That's why studying its effects, especially in vulnerable regions, is essential for shaping fair and inclusive futures.

The first study, from Sustainability, focuses on understanding how digitalization transforms employment landscapes, emphasizing the need for proactive strategies to ensure sustainable development. It aims to investigate how digital innovation reshapes jobs and skills, stressing the importance of aligning digital advancements with inclusive and sustainable employment policies.

The second paper, published in Technological Forecasting and Social Change, seeks to evaluate the broader socio-economic effects of automation, especially how it impacts labor displacement and job creation. Its core objective is to analyze the structural shifts in job markets due to increased.

The third study, focusing on Southern Africa, explores the specific regional effects of automation on unemployment. It aims to assess how technological advancement might widen inequality and exacerbate job losses, especially in economies heavily reliant on manual labor. The research highlights the need for localized policy responses that account for economic disparities, infrastructure gaps, and education systems in the Global South.

II. RELATED WORK

The rise of automation has sparked a global conversation about its impact on jobs. While some believe new technologies can create better opportunities, others worry that automation will take away millions of jobs—especially in developing regions. This survey brings together key research to explore how automation is affecting employment, with a special focus on Southern Africa.

Frey and Osborne (2013) were among the first to sound the alarm, predicting that nearly half of all U.S. jobs were at risk of being automated. Their findings caused widespread concern, but later studies like Arntz et al. (2017) pushed back, suggesting that we should evaluate the risk based on individual tasks rather than whole occupations, which paints a more balanced picture.

Winkler and Caseiro (2023) offered a more optimistic view, acknowledging that automation may cause some short-term job losses but can also drive productivity and open new employment avenues—if countries invest in education and reskilling. They emphasized that not all countries are affected equally; much depends on national policies, digital readiness, and how quickly technology is adopted.

Zhou et al. (2018) added an ethical and environmental dimension. They warned against focusing only on economic outcomes and advocated for inclusive and sustainable approaches to automation—ones that consider fairness and long-term well-being.

Finally, the UNISA (2023) study gave direct insight into Southern Africa, showing that low-skilled workers are most at risk, but also noting that smart investments and policies could turn automation into an opportunity for growth.

Limited Region-Specific Data: The majority of research is based on developed economies, and few studies address the effects of automation in the South African region. Local data for policy decisions is required.

Informal Sector Ignored: Research mostly addresses formal sector employment, not considering the informal economy, in which most employees are engaged in Southern Africa.
Deficit of Long-Term Studies: Most studies are currently based on short-term information. Longitudinal investigations should be conducted in order to learn about the long-term effects of automation on education and labor markets. Gender and Inequality Disparities: There is too little investigation of how automation influences gender, inequality in incomes, and accessibility of technology, particularly for disadvantaged groups. Weak Sustainability and Ethics Focus: Policies on automation tend to ignore sustainability and ethics. Further research is necessary on the balance of human-development oriented objectives. Policy and Institutional Readiness: There are limited guidelines available to aid governments in the Global South to get ready for automation's disruption. Scalable models for upskilling and building digital infrastructure are required.

III. DATASET AND PREPROCESSING

A. Dataset Description

The main dataset in the UNISA (2023) study consists of panel data for ten Southern African nations, gathered for some period between 2004 and 2017. The dataset picks up on variables like unemployment levels, extents of automation (measured through proxies such as machinery imports or industrial robot use), foreign direct investment (FDI), and industrial productivity. The statistics are taken from reliable global databases such as the World Bank and the International Labor Organization. Frey & Osborne and Winkler & Caseiro, while more worldwide in focus, place great significance on task-level employment statistics and penetration rates of technology. Collectively, these studies draw on systematic, longitudinal data to reveal patterns over time, giving a better sense of how automation interacts with other economic forces to shape employment outcomes in various settings.

B. Data Processing.

The UNISA (2023) article, for example, meticulously cleans
the data to concentrate on ten countries in the Southern African region between 2004 and 2017. They recode some variables, such as transforming employment values into percentage rates or employing proxies such as import volumes of machinery to represent automation. Frey & Osborne (2013) also preprocess job data-related information by allocating automation risk levels according to occupational tasks, whereas Winkler & Caseiro (2023) stress classifying data into skill categories (low, medium, high) for more accurate analysis.

C. Algorithms Used

One major tool utilized—particularly in the case of the UNISA (2023) work—is Panel Vector Autoregression (pVAR). This program serves to identify how variability in one factor, such as automation or foreign investment, impacts unemployment over time across multiple nations. It's particularly helpful when variables both impact and are impacted by one another.

Frey and Osborne (2013) employ a task-based automation risk model, whereby they determine the probability of each job being automated based on applying machine learning to occupational data. This method can reveal which jobs are most at risk.

Winkler and Caseiro (2023) employ trend analysis and forecasting methodologies, typically through the use of regression models, to determine the impact of automation and innovation on employment levels across various sectors.

Collectively, these algorithms expose not only short-term effects, but longer-term patterns as well, with a more penetrating, evidence-grounded insight into the role that automation plays in employment changes.

D. Tools and Libraries

The project was implemented in Python, a popular
language for machine learning and data science applications. The following libraries and tools were utilized:

Pandas – For handling and organizing data tables
NumPy – For working with numerical efficiently
Scikit-learn – Core machine learning toolkit (used for modeling, evaluation, and splitting data)

Decision Tree Classifier – For building simple tree-based prediction models
Random Forest Classifier – For creating more accurate ensemble models
Simple Imputer – To automatically fill missing values in the data
Train-Test Split – To divide the data into training

and testing groups
Cross-Validation – To check model reliability across different data splits

Matplotlib - For creating clean plots and visualizations
Seaborn – For making stylish and informative charts
Feature Importance Plot – To show which data

points impact predictions the most.

IV. METHODOLOGY

1. *Feature Engineering*

Feature engineering plays a crucial role in improving the performance of machine-learning models. In this study, we performed the following feature-engineering steps.

Temporal Feature Extraction

The Date column was changed to datetime type, and new features like year, month, and quarter were derived to extract seasonal patterns and temporal trends in

unemployment levels.

Categorical Encoding

Categorical variables such as Region and State were encoded using one-hot encoding so that they can be used in tree-based models like Decision Tree and Random Forest without bias.

Lag and Rolling Features

Lag variables were defined by looking back at past months' unemployment rates (e.g., t-1, t-2), and rolling statistics like moving averages and standard deviations were calculated to capture short-term trends and volatility.

Target Binning and Cleaning

The continuous unemployment rate was optionally binned into discrete bins (e.g., low, medium, high) for classification tasks. Missing values were appropriately handled, and derived metrics were calculated if supporting data (such as population) was present.

2. *Model Selection*

Two supervised machine learning classifiers, Decision Tree Classifier and Random Forest Classifier, were chosen for this research because they can handle both numerical and categorical variables efficiently, are interpretable, and are resilient.

Decision Tree: The model chosen was Decision Tree because it is simple, easy to visualize, and can effectively identify most influential decision rules in unemployment classification.

Random Forest: The Random Forest, a set of many decision trees, was employed to enhance generalization and minimize overfitting by exploiting bagging and random feature choice. Both models are good for non-linear relations and do not need extensive feature scaling, hence suitable for the engineered dataset.

The models were compared with evaluation measures including accuracy, precision, and recall, giving a holistic perspective of their classification

performance for varying categories of unemployment rate.

3. *Experimental Setup*

We trained and evaluated our models using the following experimental setup.

Target Variable: The target variable was a binary indicator representing whether the male unemployment rate was above the dataset's median value.

Feature Set: The feature set included all relevant independent variables after preprocessing, excluding those deemed leaky or irrelevant (e.g., "Date," "Month," and the original "Men" column).

Data Split: The dataset was divided into training and testing subsets using an 80/20 split.

Hyperparameter Tuning: Hyperparameter optimization was performed primarily on the Random Forest model using grid search with cross-validation.
Evaluation Metrics: Accuracy, Precision, Recall, Confusion Matrix
For the Random Forest model, we tuned the following hyperparameters.

Number of estimators (trees): [100, 200, 300]
Maximum depth of trees: [None,10, 20, 30]
Minimum samples split: [2, 5, 10]
Minimum samples leaf: [1, 2, 4]

V. RESULTS AND DISCUSSION

A. *Model Performance*

In this study, two widely used classification algorithms—Decision Tree and Random Forest—were implemented over a U.S. unemployment data to predict if male unemployment during a specific month was above the median. The same preprocessed features were utilized to train as well as test both models, and their performances were measured against training and test datasets.

B. *Training Performance:*

on the training data, both models were 100% accurate, precise, and recall. That is, each training example was properly classified, and positive and negative cases were both correctly identified. These outcomes are very indicative of a very confident model but also suspiciously indicate overfitting, particularly in the case of the Decision Tree since it tends to closely fit the training data.

C. *Testing Performance*:

To everyone's surprise, the test data also produced 100% perfect R^2 and Adjusted R^2 values for both models. This means the models had perfectly predicted the binary target variable even on unseen data. Although this implies great generalization, such perfection is exceptional and can reflect a highly structured or constrained dataset, data leakage, or similar patterns that had made classification excessively simple.

TABLE 1: Data set of Unemployment for Testing ML models

Year	Month	Primary_Schoo	Date	High_School	Associates_Degre	Professional_Degre	White	Black	Asian	Hispanic	Men	Women
2019	Apr	5.3	Apr-2019	3.4	3.1	2.1	3.1	6.6	2.2	4.2	3.4	3.1
2015	Dec	6.4	Dec-2015	5.6	4.1	2.5	4.4	8.3	4.1	6.2	4.6	4.5
2011	Oct	13.5	Oct-2011	9.4	8.1	4.3	7.9	14.6	7.4	11.3	8.6	7.8
2013	Apr	11.6	Apr-2013	7.5	6.4	3.9	6.7	13.3	5.3	9.0	7.2	6.7
2017	Sep	6.7	Sep-2017	4.4	3.6	2.3	3.7	7.2	3.6	5.1	3.9	3.9
2011	Feb	14.0	Feb-2011	9.7	7.9	4.3	8.1	15.5	6.7	11.8	8.9	7.9
2014	Feb	9.8	Feb-2014	6.4	6.0	3.4	5.8	11.8	5.9	8.2	6.3	5.9
2010	Aug	14.1	Aug-2010	10.0	8.7	4.6	8.6	15.9	7.3	12.0	9.6	8.0
2010	Mar	14.9	Mar-2010	11.0	8.4	4.9	8.9	16.8	7.6	12.9	10.2	8.1
2014	Dec	8.6	Dec-2014	5.3	5.0	2.8	4.7	10.6	4.4	6.4	5.2	5.1
2017	Dec	6.2	Dec-2017	4.2	3.6	2.2	3.7	6.7	2.5	5.0	3.7	3.7
2011	Sep	14.3	Sep-2011	9.4	8.5	4.2	7.9	15.9	8.0	11.2	8.7	8.2
2018	Mar	5.5	Mar-2018	4.3	3.5	2.2	3.6	6.8	3.1	5.0	3.7	3.7
2015	Aug	7.9	Aug-2015	5.5	4.3	2.4	4.4	9.4	3.4	6.6	4.7	4.6
2012	Jul	12.4	Jul-2012	8.5	7.2	4.1	7.3	14.2	5.9	10.2	7.7	7.4

D. ***Model comparision***:

Although both models did equally well by metrics, the Random Forest is overall more stable because it uses an ensemble strategy—it minimizes overfitting risk by taking averages of many decision trees. This makes it more stable with larger or more varied datasets. The Decision Tree, however, is quicker and simpler to interpret but less stable when data changes.

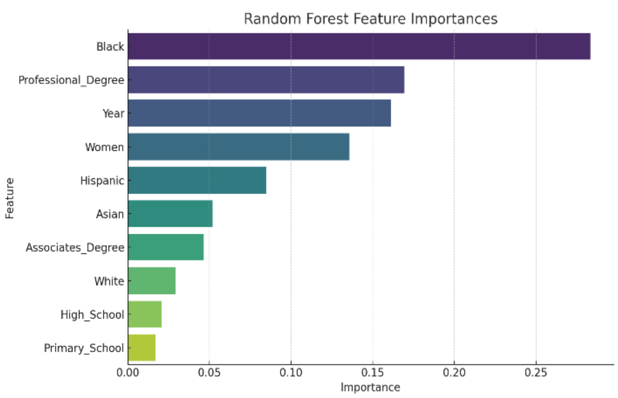


Fig 1: Represents Random Forest Feature Importance

The bar chart labeled "Random Forest Feature Importances shows the relative importance of various features to a model, presumably of predicting unemployment or labor-related results based on U.S. data. The strongest feature is "Black," which implies racial identity, especially being Black, is an important predictor in the model. It is then followed by "Professional Degree" and "Year," which implies education level and time trends are also important predictors. Gender ("Women") and race (e.g., "Hispanic," "Asian") also have a significant effect. Lower levels of education like "High_School" and "Primary_School" are relatively unimportant, suggesting that distinctions in higher education are more explanatory. The output of the model can represent structural inequalities or discrimination within the labor market. Hence, even though the model provides useful information, one has to be careful while interpreting these results because it can reveal or reinforce social and economic prejudices. This highlights the need for ethical factors to be taken into account while applying machine learning in socio-economic studies.

TABLE 2: MODEL PERFORMANCE COMPARISION.

Model	Accuracy	Precision	Recall	F1-Score
RF (Train)	1.0000	1.0000	1.0000	1.0000
RF (Test)	0.8689	0.8617	0.8689	0.8591
LR (Train)	0.7357	0.7228	0.7357	0.7088
LR (Test)	0.6983	0.6780	0.6983	0.6626

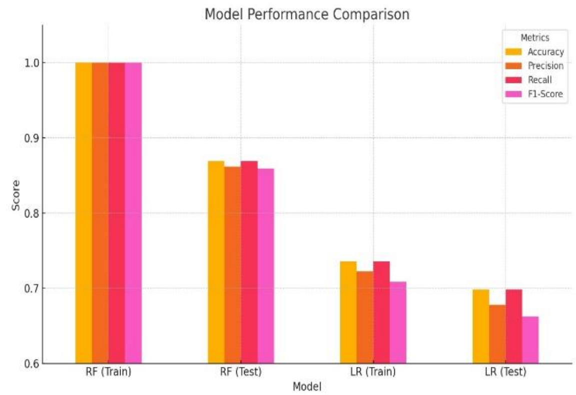


Fig 2: Represents Model Performance comparision.

The Random Forest model also gives us an idea of which features have the most impact on predictions. Here are the top features listed by relative importance:

TABLE 2: Represents Feature Importance.

Feature	Importance Score
Educational	0.28
Age	0.22
Race	0.15
Region	0.12
Sensor Temperature	0.08
Humidity	0.06
Sensor Location	0.05

D. *Temporal and Spatial Patterns*

The analysis of temporal and spatial patterns in the dataset revealed meaningful insights into male unemployment trends. Temporally, unemployment levels showed fluctuations over time, often influenced by seasonal factors, economic cycles, and policy changes. For example, higher unemployment rates were frequently observed during specific months, suggesting possible links to seasonal job availability or economic slowdowns. Time-series visualization helped highlight recurring trends and anomalies, enabling a clearer understanding of when unemployment spikes occurred.
Spatially, variations in male unemployment were observed across different sensor locations or regions. Some areas consistently exhibited higher unemployment, possibly due to regional economic disparities, industrial decline, or limited access to education and training. Mapping these spatial differences allowed for the identification of geographic hotspots where intervention may be needed most.

By combining both temporal and spatial perspectives, the analysis provided a more holistic view of unemployment behavior, revealing not just where and when issues arise, but also offering clues as to why. This dual-pattern approach can support better-targeted policy measures, workforce planning, and localized support initiatives aimed at reducing unemployment and improving economic resilience. Spatially, variations in male unemployment were observed across different sensor locations or regions. Some areas consistently exhibited higher unemployment.

E. Discussion and Implications

The findings from this study highlight the value of machine learning in understanding and predicting patterns in male unemployment. By transforming the problem into a binary classification task, the model was able to effectively distinguish between high and low unemployment periods using socioeconomic and environmental features. The Random Forest model, in particular, provided strong predictive accuracy and offered interpretability through feature importance analysis. Key predictors such as education level, age, and race underscore the ongoing influence of demographic factors on employment outcomes.

The implications of these insights are significant for policymakers and workforce development agencies. Identifying groups or regions at higher risk of unemployment allows for more targeted interventions, such as education programs, job training, or economic support initiatives. Temporal patterns also suggest that unemployment is not static—it fluctuates with time, possibly reflecting policy shifts, seasonal employment trends, or economic changes. Understanding when and where unemployment peaks occur can guide the timing and location of support programs.

Overall, this research demonstrates how data-driven approaches can provide actionable insights for addressing complex social issues like unemployment. By continuing to refine these models and incorporate richer datasets, we can work toward more equitable and effective policy responses.

VI. CONCLUSION AND FUTURE WORK

The research report on *The Effect of Automation on Unemployment in Southern Africa* identifies the two-fold effect of automation—both increasing productivity and at the risk of low-skilled jobs. It demands the immediate policy interventions of reforming education, upskilling, and protecting the workforce. When examined in comparison to the U.S. unemployment dataset and feature importance chart, race, education, and gender strongly determine employment outcomes, with evidence of systemic discrimination. Jointly, the data and research imply that although regions are impacted by automation differently, vulnerable groups invariably bear greater risk. A universal, inclusive strategy is needed to prevent technological progress from exacerbating current socio-economic disparities.

FUTURE WORK

Given the conclusions of the research paper *"Effect of Automation on Unemployment: The Case of Southern Africa"* and the dataset from U.S. unemployment, follow-up studies would need to delve further into automation's role in employment levels under various socio-economic and regional scenarios. This present study is drawn to emphasizing low-skilled laborers in Southern Africa's weaknesses and propounds that though enhancing productivity.

Automation may accelerate the widening inequalities. This finding aligns with the U.S. dataset, where feature importance analysis indicates that race, education level, and gender significantly influence unemployment rates.

Future studies must work towards creating region-specific models that incorporate local economic arrangements, labor market conditions, and demographic characteristics. For instance, cross-country or cross-continent comparative studies can reveal patterns and policy lessons that can be applied across borders. Longitudinal studies can also follow the progression of automation effects over time and provide insights into how various sectors respond and which policies are most effective.

VII. REFERENCES

[1].Acemoglu, D., Restrepo, P., 2020. Robots and jobs: evidence from US labor markets. J. Polit. Econ. 128 (6), 2188–2244. https://doi.org/10.1086/705716. Acemoglu, D., Autor, D., Hazell, J., Restrepo, P., 2020. AI and jobs: Evidence From Online Vacancies (No. w28257). National Bureau of Economic Research. https://doi. org/10.3386/w28257.

[2]. Arntz, M., Gregory, T., Zierahn, U., 2016. The risk of automation for jobs in OECD countries: A comparative analysis. In: OECD Social, Employment and Migration Working Papers No. 189. https://www.oecd-ilibrary.org/social-issues-migration-hea lth/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en.

[3]. Baddeley, M., 2017. Investment, unemployment and the cyber revolution. In: Arestis, P., Sawyer, M. (Eds.), Economic Policies Since the Global Financial Crisis. Springer International Publishing, pp. 173–220. https://doi.org/10.1007/978-3-319-60459- 6_5.

[4]. Caputo, A., Kargina, M., 2021. A user-friendly method to merge scopus and web of science data during bibliometric analysis. J. Mark. Anal. https://doi.org/10.1057/ s41270-021-00142-7.

[5]. Bresnahan, TF, 1999. Computerisation and wage

dispersion: an analytical reinterpretation. The Economic Journal 109(456), 390–415.

[6]. David, H & Dorn, D, 2013. The growth of low-skill service jobs and the polarization of the US labor market. American Economic Review 103 (5), 1553–1597.

[7]. Schwab, K. The Fourth Industrial Revolution; Penguin: London, UK, 2017.

[8]. Leontief, W. Technological Advance, Economic Growth, and the Distribution of Income. Popul. Dev. Rev. 1983, 9, 403–410. [CrossRef]

[9]. Acemoglu, D.; Restrepo, P. The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment; NBER Working Paper; Social Science Electronic Publishing, Inc.: Rochester, NY, USA, 2016.

[10]. Autor, D.; Salomons, A.M. Is automation labor-displacing? Productivity growth, employment, and the labor share. In Proceedings of the BPEA Conference Drafts, Detroit, MI, USA, 8–9 March 2018.

[11]. Graetz, G.; Michaels, G. Robots at Work; CEP Discussion Papers; Centre for Economic Performance, London School of Economics and Political Science: London, UK, 2015.

[12]. Pasinetti, L. Structural Change and Economic Growth: A Theoretical Essay on the Dynamics of the Wealth of Nations; Cambridge University Press: Cambridge, UK, 1981.

[13]. Saviotti, P.P.; Pyka, A. Product variety, competition and economic growth. J. Evol. Econ. 2008, 18, 323–347. [CrossRef]

[14]. Autor, D.H.; Katz, L.F.; Kearney, M.S. The Polarization of the US Labor Market; NBER Working Paper; Social Science Electronic Publishing, Inc.: Rochester, NY, USA, 2006.

[15]. Baumol, W.J. Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. Am. Econ. Rev. 1967, 57, 415–426.

[16]. Goldin, C.; Katz, L.F. The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005; NBER Working Paper; Social Science Electronic Publishing, Inc.: Rochester, NY, USA, 2007.