JOB TRANSITION: A CASE OF MITIGATION AGAINST AUTOMATION?

PHILIPPE PANHALEUX, ARYAZ EGHBALI, ROGER WATTENHOFER

1. Introduction

The global economy has changed forever with the rise of computers. Through smaller, cheaper, and faster components, a new wave of advances in artificial intelligence (AI) and machine learning are bringing about fundamental changes in the workplace. IBM's Watson and DeepMind's software now regularly solve problems previously thought to be impossible to automate, often more successfully than humans (Ferrucci D., et al. 2010; Silver D., et al. 2017). Mechanical robots have also become increasingly advanced, with many high-end cars now featuring some level of autonomous driving or robots being used to transport items in warehouses (D'Andrea R., et al. 2012). Due to these recent innovations, concerns have arisen across the world that automation would destroy many jobs – even non-routine ones – leading to widespread joblessness. This fear bears some similarities with that voiced during the Industrial Revolution. To prevent computers from putting thousands of people out of work, many initiatives have been laid down to retrain workers. Adopting the same approach as that used in the Industrial Revolution period, when the workforce learnt to operate machines, it seems that now we should move to operating and designing AI and its code. Amazon has decided to teach software engineering and IT support to many of its warehouse workers through Associate2Tech programmes (Matsakis L. 2019). As coal plants are to be phased out, many coal miners will lose employment. To help them get a new job, the Mined Minds program has started to train them to become programmers (Jiang W. 2017).

However, many job retraining programs have not been particularly successful, with some of them being heavily criticised (Robertson C.

2019; D’Souza S. 2019). While new software engineers are in high demand in today’s world of work, it seems a bit naive to believe that every truck driver or miner can easily turn into a programmer twenty years after leaving high school. It seems far more realistic to look at which jobs will still be needed in the future, and then look for those which are closer to the occupations facing a high-risk of automation. One would expect such a transition to be far less difficult.

It is in this context that we analysed the potential for computerisation, by modelling the probability of automation for most occupations in the United States, pointing out correlations between automation and job characteristics. We then found potential replacement occupations for threatened workers and looked at whether these occupations would see an increase in demand over the next decade, hoping that they would be able to cope with the influx of new workers.

2. Related Work

Much research has been published on automation in the last decade, due to growing concerns about unemployment and massive changes in the labour market (Arntz M., Gregory T., Zierahn U. 2016; Arntz M., Gregory T., Zierahn U. 2017; Nedelkoska L., G. Quintini G. 2018; Wike R., Stokes B. 2018).

One of the most famous papers is that by Frey and Osborne (2013). Using a database compiled by the U.S. government and some machine learning methods, they obtained the probability of automation for over 700 occupations, classifying them as having a low, medium, or high risk of being automatized. The authors found that 47% of workers were employed in high-risk occupations, drawing global attention to the possible threat caused by what they called “computerisation”.

Following that study, Brandes and Wattenhofer studied the issue in more detail (2016). They took a more granular approach and determined the likelihood of automation for each task within a single occupation, using Frey and Osborne’s results. They classify each occupation according to their respective tasks, defined a quantifiable frequency at which these tasks are carried out, and computed the probability of automation for all tasks.

The McKinsey Institute also published a report about job automation and its impacts (Manyika J., et al. 2017). Although this report is a thorough study in this area and considers multiple countries, it does not explain how to tackle this problem. Another major study related, albeit tangent, to our questions was carried out by Bakhshi et al. (2016). Their study provides predictions for the future opportunities in each occupation. It was their results that allowed us to find appropriate jobs workers can switch to.

3. The Future of Jobs

After noticing biases and some shortcomings in Frey and Osborne (2013), we decided to compute the probability of automation once again instead of drawing on previous research results. Using a more detailed and objective methodology, we obtained a more consistent and general result which is not based on subjective analysis of the data and the problem at hand.

3.1. Data

Just like the previously mentioned studies, we used the O*NET database of the Bureau of Labor Statistics (BLS). We considered the latest version at the time of writing, namely that of February 2019[1]. The O*NET database is a convenient tool to analyse the properties of jobs, as it uses the Standard Occupational Classification (SOC) to define occupations and provides many different variables to describe them. Its data is based on the results of surveys collected from professionals in every field.

While the data presented in O*NET is potentially interesting, we decided to restrict our research to a few types of variables. The most important ones are requirements, work contexts, and work activities. There are three types of requirements: skills, abilities, and knowledge. They contain 120 variables which describe what a worker needs to know or do to be effective at work. The work context of an occupation describes the environment in which the worker is active, and the work activities are a very general description of what the worker actually does.

All data in O*NET is classified using a numerical scale. Work contexts, which indicate the working environment and conditions, simply range from 1 to 5 (or 1 to 3 in two specific cases). Work activities and requirements, however, are actually described using two different scales: importance and level. ‘Importance’, going from 1 to 5, describes how prevalent this activity or this requirement is during the job. ‘Level’, ranging from 1 to 7, describes how complex the requirement or activity is (simply walking would be 1, whereas piloting a spacecraft to the moon would be 7). We used both variables by simply multiplying one by the

[1] Other papers used data from 2010, which could lead to differences in the characteristics and definitions of occupations.

other for every variable and occupation, resulting in a metric showing how valuable the requirement or activity is. To compare our results with wage and employment statistics, we had to select the occupations which are also present in other BLS datasets. One issue is that O*NET sometimes further divides occupations into separate, even more specialised jobs. In this case, we averaged the values of each variable of these different "sub-occupations".

3.2. Modelling

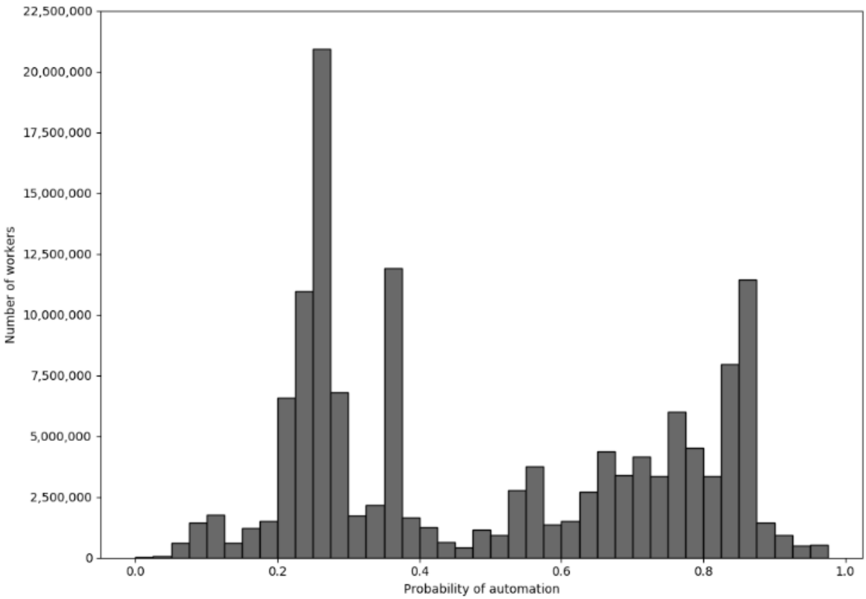
In order to determine the automation probability using our data, we need a regression model which can extract the results by looking at every variable and noticing patterns in the dataset. While neural networks have become almost ubiquitous in modern machine learning, they are not adapted to small datasets like ours, whereas linear regression would be too simple. An excellent method which is both flexible and adapted to our needs is the Gaussian process. It is essentially a stochastic process whose variables follow a multivariate, normal distribution. Like any regression task, we need some training data for the prediction function. We labelled 70 occupations out of 749 with a probability of either 0 or 1, depending on the answer to the following question: "What is the likelihood that these occupations will be automated sometime in the future?". We logically only labelled the occupations providing a sure answer. We also made sure that the labels were well distributed among every major occupational group, so that the results are not skewed by any specific job type. Finally, there is an equal amount of 0 and 1 labels, as we did not want to give more weight to one end of the scale than to the other. We compared our labels with those of Frey and Osborne, and looked at major trends in automation, machine learning, and robotics, to answer the question.

The issue when using every variable in three distinct models is twofold. On the one hand, too many variables will muddle the model and make it noisier. Most variables do not have a particularly important impact on output, and as such can be discarded for a more certain result without influencing the result too much. On the other hand, looking at either requirements, work activities, or work contexts separately may disregard the covariance between the variables of different groups. The best way to reconcile both issues would be to produce a single, big model, and only afterwards discard the irrelevant variables. With a total of 240 variables but only 70 labelled data points, however, the model would be too ill-conditioned, and computing its output becomes virtually impossible. We therefore decided to model the three different outputs, and discard the useless variables in each of them, before putting the relevant ones together for a last pass through the regression method.

There are many different ways to do a so-called feature selection. One of the easiest methods is to compute a Pearson correlation coefficient for each variable, but its linearity makes it a poor choice in our case. A rather common technique found in machine learning is Principal Component Analysis (PCA), but it is not without shortcomings either. More modern techniques such as explanation vectors, LIME, and Shapley values offer a detailed and versatile way to understand how black-box models have reached their conclusion. Instead of comparing all of these methods and choosing the best one, we employed SHAP values. They combine both LIME and Shapley values for a more model-agnostic and interpretable result.

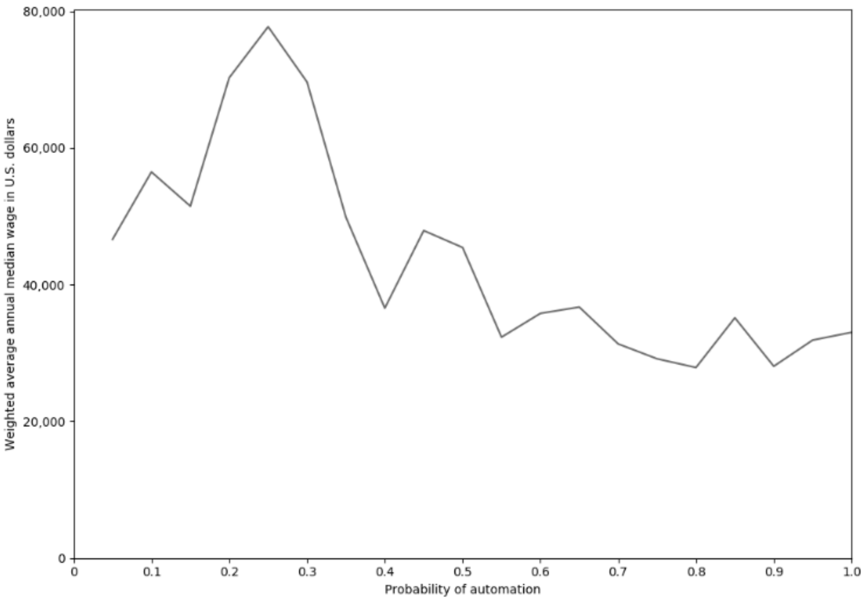
3.3. Results

Figure 1: Distribution of the probability of automation for U.S. workers.



After computing our predictions for the three separate classes of variables, we used the SHAP library (Lundberg S.M., Lee S.-I. 2017) to select the most relevant features. We chose 6 requirements, 7 work activities, and 7 work contexts. Using these 20 variables, we predicted the likelihood of computerisation one last time. We then added the respective employment statistics to every job to evaluate the amount of workers at risk of automation.

Figure 2: Annual median wage of U.S. workers against probability of automation.



We found that 32% of workers are employed in so-called high-risk occupations, which are defined as having a probability of automation of over 70% (Figure 1). About 30% are at medium risk of computerisation (the probability is between 30% and 70%), and the remaining 38% are at low risk of being automatized. These are slightly more optimistic numbers than Frey and Osborne's. The reader can also notice that many more jobs have uncertain prospects. This is due to our less categorical modelling, as more variables increase uncertainty. For example, railroad conductors have a probability of automation of 75.3%, while ship engineers stand at 8%.

We also noticed that occupations with a higher annual median wage are on average far less likely to be automated than low-paying occupations, with one exception being trade occupations (Figure 2). Surgeons have a probability of automation of 26% while the likelihood that fast food cooks will be replaced by robots stands at 83%.

Finally, we can see on Figures 3 and 4 that most low-risk occupations require a far higher level of education than high-risk occupations, and also require more work experience. Again, trades are a noteworthy exception to the main trend. For example, chief executives have a probability of automation of 23.3%, while Heavy and Tractor-Trailer Truck drivers report a probability of 82.0%. Plumbers have a 15% of being ousted by robots.

Figure 3: Required education level against probability of automation.

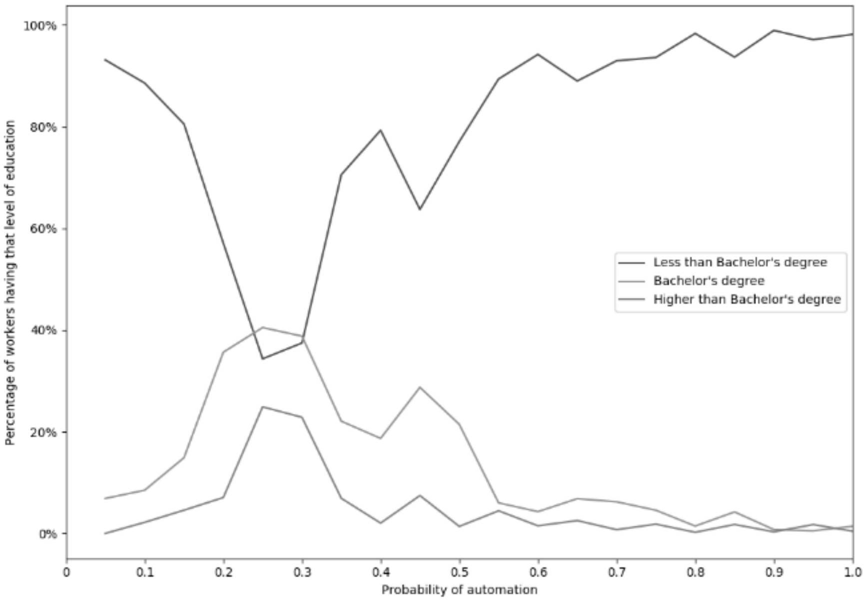
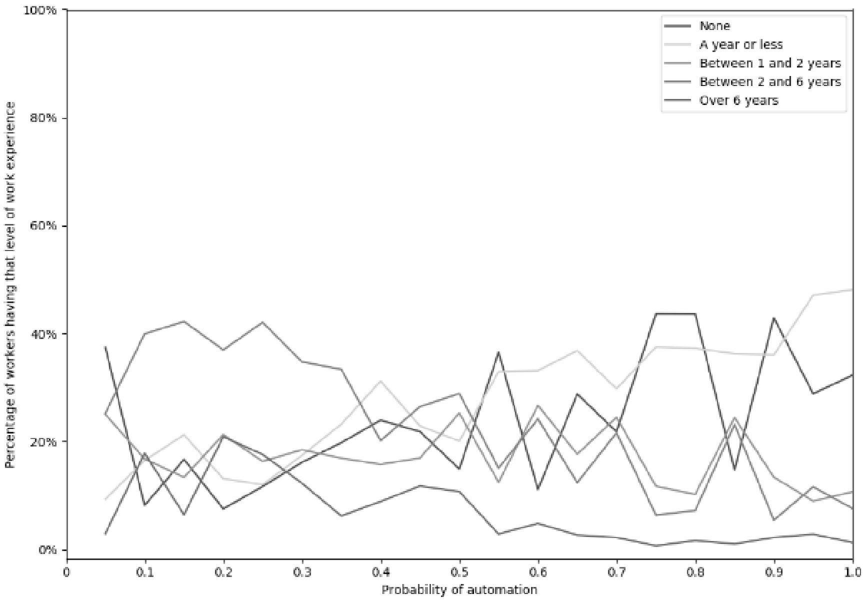


Figure 4: Required work experience against probability of automation.



Now that we have the data about which occupations are at risk, and which are not, we can compare them to one another and see which low-risk jobs offer a transition opportunity to high-risk workers.

4.1. Methodology

First, we must define similar occupations. O*NET provides several levels of variables to accurately describe the activities done within an occupation. Tasks are the lowest level and have a unique identifier and description. Therefore, there are no two jobs having the same task. The next level is the Detailed Work Activity (DWA). Every task is linked to one or more DWAs. We can therefore define which of these are assigned to which occupations. As DWAs have a standardised identifier, we can compare occupations using those.

We now need to define the notion of a 'closely related occupation' using DWAs. It is clear that a specific percentage of common activities is required. We decided to consider the percentage of work activities of the low-risk occupation which can be also found in the high-risk occupation.

It is more interesting to know how familiar a worker would be with a potential new occupation than with their old, high-risk one. We are faced with two issues. First, the cut-off has to be quite arbitrary, and second, this system does not take into account task frequency (as opposed to Brandes and Wattenhofer's). For the first problem, we decided to take a 25%-similarity threshold. The transition outlook will obviously look better with a lower value and worse with a higher one. The reasoning for having chosen 25% is that the value is high enough to allow a worker to be at least partly productive right out of the gate, and low enough to allow for some inter-occupation variation, since not every job is the same. One also has to consider that DWAs are very detailed, so different DWAs may not be fundamentally different. We decided not to take into account frequency, believing that most tasks take up a similar portion of the job.

4.2. Demand Prospects

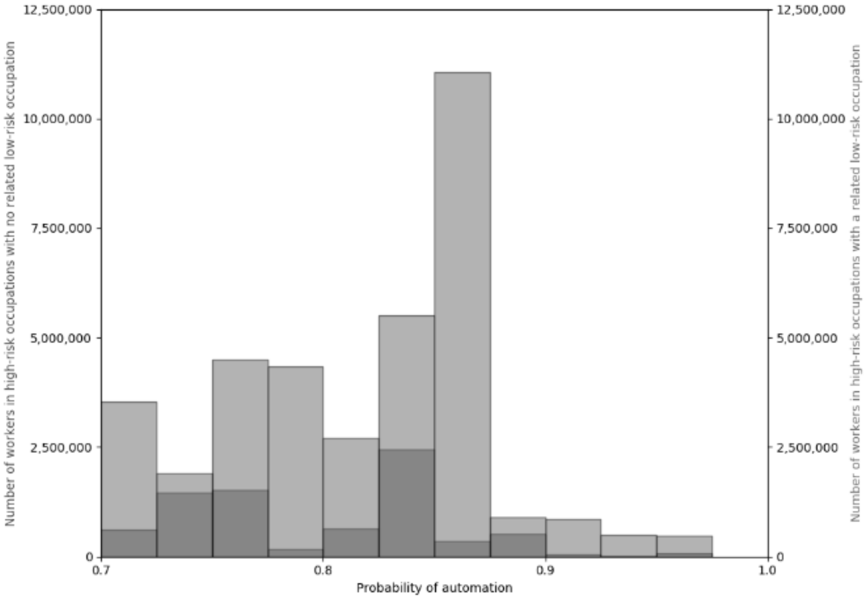
Knowing whether someone can move from a high-risk job to a low-risk one is a very important step in understanding the occupational outlook of vulnerable workers. However, our first analysis will only show whether a transition is theoretically possible. After all, the economy may not need more workers in a low-risk occupation closely related to a high-risk one. We therefore also need to factor in the demand-increase probability for every low-risk occupation.

Using O*NET data and all 120 requirements, Bakhshi et al. modelled the probability of automation using a heteroskedastic ordinal Gaussian regression. Two workshops were held, each one of which was attended by some twelve experts in various fields. The labels were defined by these experts, which had to determine if the absolute number of workers in a given occupation would increase, decrease, or remain stable, and if the number of workers relative to the total employed population would increase, decrease, or remain stable. They answered those questions for 10 different occupations, after which an active learning algorithm selected 10 new occupations to be labelled. In total, 30 occupations were assigned labels. The experts were also asked to rate how certain they were, which also influenced model output. The authors computed a probability of increase in demand between 0.0 and 1.0 for 747 different occupations.

4.3. Results

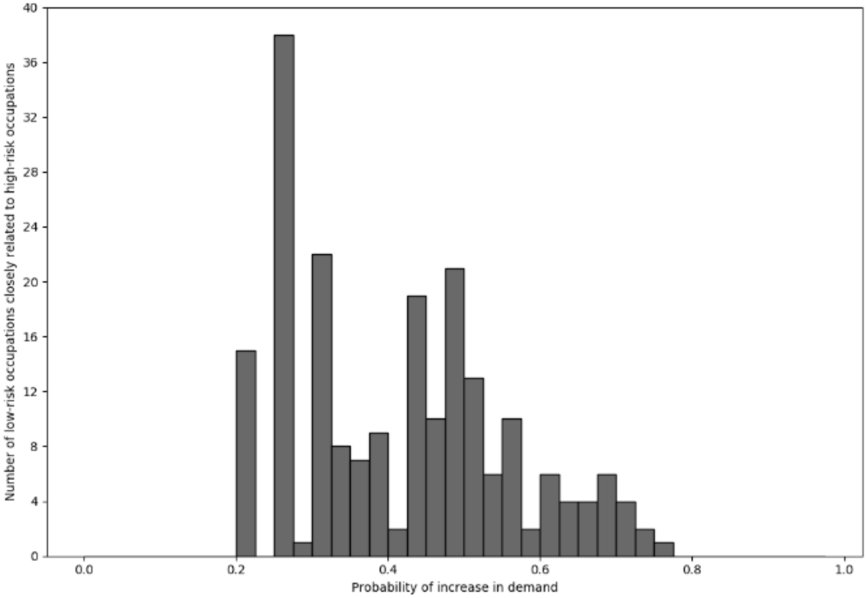
Now that we know how to obtain our high-risk/low-risk related occupations pairs and the probability of a demand increase for each of the low-risk occupations, we can finally determine whether workers at high-risk of automation can transition to a low-risk occupation, and whether the demand for these low-risk jobs is likely to increase.

Figure 5: Number of workers at high-risk of automation and their potential to transition to low-risk jobs.



We note that there are 182 occupations at high risk of automation (Figure 5). 117 of those have at least one closely related occupation with low potential for computerisation, whereas 65 of them do not have a closely related occupation with a low probability of automation. Surprisingly, occupations that do have closely related low-risk occupations actually tend to have a higher probability of automation. While occupations with no close relations represent a third of all high-risk occupations, they actually employ a far lower share of workers at high-risk of automation. This indicates that many of these workers could potentially move over to similar jobs with a low automation probability. For example, railroad conductors, which, as mentioned previously, are at high-risk of automation, have a job which is related to that of ship engineers, who are at low risk.

Figure 6: Probability of increase in demand for low-risk occupations related to high-risk ones.



We now only select the 117 occupations which have a closely related low-risk occupation. While we previously saw that many workers could move to low-risk occupations, Figure 6 provides a bleaker picture. Most low-risk closely related occupations do not have particularly good demand prospects, with the vast majority having a rather uncertain outlook and more than half being more likely to decrease than increase in demand: only 39 of the occupations have a closely related occupation with a probability of increase in demand higher than 0.5. To continue with our example, ship engineers may be a possible occupation for railroad conductors to transition to, but the probability that demand for ship engineers will increase in the future only stands at 33%.

5. Conclusion

Our results show that most workers at high risk of automation can indeed transition to safer jobs which are at least partially similar to their current one. We observed which values were most important instead of selecting them ourselves, providing a more data-oriented approach. Most jobs to transition to do have a higher required level of education, which implies that some training will be needed. Finally, it is not clear whether retraining will actually be worth it. Many safer jobs have uncertain demand prospects and may therefore not be able to accommodate so many new workers. We hope that our results can show which steps are needed to enable workers to smoothly transition to new occupations and adapt to the automation of their current occupation. We obtained valuable results with a simple model. However, a more complex model might be able to capture some job-specific aspects better. For example, we used the same fraction for all jobs to detect similarities between them. An in-depth investigation can label parts of each job to their similarity degree, which would result in a more expressive model. Nonetheless, predicting the future, especially about the evolution of technology and its impact, is, and continues to be, a difficult exercise.

5.1. Limitations

The labelling process remains the weakest link of any quantitative effort to analyse the impact of automation. The estimation of which occupations are surely automatable, and which are not, will always feature a degree of uncertainty and subjectivity. We do not see a way to remedy that problem, except analysing every occupation in detail, which would require a large and sustained effort of many experts to accurately predict the future of each occupation.

5.2. Future work

While we are confident that our results lead to a better understanding of the situation, we see much potential for more research on the subject. One big aspect is location. By looking at geographical variations in the automation risk and the education potential in high-risk areas, one could estimate how complex and socially disturbing retraining could be. The costs and benefits of such programs should also be analysed, and their potential compared to existing job retraining initiatives. Finally, many jobs existing today did not exist twenty years ago. It may be wise to study more closely why and how these occupations appear, their characteristics, and how current workers can be trained to find a job in those emerging sectors.

References

Ferrucci D., et al. (2010). Building Watson: An Overview of the DeepQA Project. *AIMag*, vol. 31, no. 3, pp. 59-79.

Silver D., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, vol. 550, no. 7676, p. 354.

D'Andrea R., et al. (2012). Method and system for transporting inventory items. *US Patent,* 8 280 547.

Matsakis L. (2019). Amazon pledges $700 million to teach its workers to code. *Wired* [Online]. Available: https://www.wired.com/story/amazon-pledges-700-million-training-workers/.

Jiang W. (2017). Out-of-work coal miners find new work in computer industry. *CBS News* [Online]. Available: https://www.cbsnews.com/news/out-of-work-coal-miners-find-new-work-in-computer-industry/.

Robertson C. (2019). They were promised coding jobs in Appalachia. Now they say it was a fraud." *New York Times* [Online]. Available: https://www.nytimes.com/2019/05/12/us/mined-minds-west-virginia-coding.html.

D'Souza S. (2019). 'They failed me': How the promise of retraining U.S. coal miners came up empty." *CBC* [Online]. Available: https://www.cbc.ca/news/world/mined-minds-retraining-appalachian-coal-miners-1.5145676.

Arntz M., Gregory T., Zierahn U. (2016). The risk of automation for jobs in OECD countries. *OECD iLibrary.* [Online]. Available: https://doi.org/10.1787/5jlz9h56dvq7-en.

Arntz M., Gregory T., Zierahn U. (2017). Revisiting the risk of automation, *Economics Letters*, vol. 159, pp.157–160.

Nedelkoska L., G. Quintini G. (2018). Automation, skills use and training. *OECD.*

Wike R., Stokes B. (2018). In advanced and emerging economies alike, worries about job automation. *Pew Research Center, Global Attitudes & Trends.*

Frey C.B., Osborne M.A. (2013). The future of employment: How susceptible are jobs to computerisation?

Brandes P., Wattenhofer R. (2016). Opening the Frey/Osborne Black Box: Which Tasks of a Job are Susceptible to Computerization? *arXiv e-prints*, p. arXiv:1604.08823.

Manyika J., et al. (2017). A Future That Works: Automation, Employment and Productivity. *McKinsey Institute.*

Hasan Bakhshi, et al. (2017). The future of skills: Employment in 2030, *Nesta* [Online]. Available: https://www.nesta.org.uk/report/the-future-of-skills-employment-in-2030/.

Lundberg S.M., Lee S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.