Measuring innovations by their labor impact

Stanford CS224N Custom Project

Wajeeha Ahmad
Department of Management Science and Engineering
Stanford University
wajeehaa@stanford.edu

Abstract

This project aims to measure the occupational impact of innovations using patent text data. Innovations can both replace tasks previously performed by human labor and augment human labor in their occupations. The extent to which a given patent automates and augments human labor is measured by comparing the textual similarity of patents with descriptions of tasks performed in various occupations and occupational titles. This project compares different techniques for obtaining such similarity measures using both static word embeddings and contextual word embeddings. I then evaluate which measures are better able to predict occupational content. This evaluation involves analyzing the extent to which automation measures for occupations obtained via static and contextualized embeddings can predict the degree to which an occupation's tasks consist of routine, abstract and manual tasks. I find that measures based on GloVe embeddings are better able to predict occupational content than contextualized embeddings.

1 Key Information to include

- Mentor: Lucia Zheng.
- External Collaborators (if you have any): N/A.
- Sharing project: No.

2 Introduction

Technological advances, particularly in artificial intelligence (AI), have led to widespread anxiety about the future of human professions. Recent findings reveal that artificial intelligence and machine learning impact most occupations [1, 2, 3] and that the AI surge is driven by task substitution whereby AI automates a subset of tasks formerly performed by human labor [4] with one estimate predicting 47% of US jobs being automated over the next decade or two [5]. While new innovations can automate tasks previously performed by labour, they can also augment human labor by creating new tasks and activities in which humans can be productively employed such as human-AI collaborative writing [6] among other tasks. Measuring such effects is challenging since the boundary between labor-augmenting and labor-automating innovations is not well-defined.

This project measures the occupational impact of innovations using patent text data. I use both static word embeddings [7] and contextual word embeddings based on BERT [8] and [9] to identify innovations that replace or create tasks performed by human labor, and match them to human occupations. I then compare how well measures generated from different embeddings can predict the task content of occupations to evaluate the different measures.

Such measures will enable us to explore whether the impact of technological change on societal outcomes such as employment growth, wages and the creation of new work varies by the type of innovation. Innovations that replace human workers via automation reduce their economic and political bargaining power whereas those that augment workers ensure that they remain integral to value creation and decision-making. Specifically, future work could examine whether innovations that automate human labor have different labor market outcomes as well as different performance outcomes for the firms that produce and use them relative to innovations that augment human labor.

3 Related Work

Previous research used simpler ways of identifying the types of innovations and technologies using specific keywords in the text of patent data to identify automation patents in machinery [10] or more generally, using the level of investment in adopting specific automation technologies such as robots. Other papers have used manual labelling of task descriptions of human occupations to identify their suitability for machine learning and exposure to artificial intelligence (Brynjolfsson et al., 2018; Felten et al., 2021). More recently, papers have used textual analysis methods to identify innovations recorded in patents that potentially overlap with the tasks performed by occupations [3, 11, 12]. A recent paper by Autor et al. (2021) is broader in its scope in identifying automation and augmentation innovations since it is not restricted to identifying specific kinds of innovations (such as machine learning, robotics or AI patents), but more generally, identifies all innovations (as measured using patent text data) that overlap with the text of occupational titles and tasks.

4 Approach

4.1 Data

This work the following three main datasets:

1. The patent text dataset consists of the full text of patents, including the summary/abstract, patent claims and patent descriptions along with numberous patent identifiers and categories.[1] I collect patent text data from patentsview.org, which provides the full text of patents, including the summary/abstract, patent claims and patent descriptions along with numerous patent identifiers and categories. For this project, I restrict my patent text data to the summary text of all patents granted in 2010 to a US firm for which a previous classification into automation and non-automation patents exists [12]. Restricting my dataset in this manner has two advantages: 1) it enables me to generate embeddings and classification scores in the time allotted for this project (using the full patent text data for all years would take prohibitively long), and 2) it enables me to generate automation scores that have an existing baseline for comparison using the classification from [12]. My sample of patents consists of 94,368 unique patents.
2. The data on occupational titles comes from the U.S. Census Alphabetical Indexes of Industries and Occupations list over 21,000 industry and 31,000 occupation titles in alphabetical order for about 700 main occupations.[2] The similarity of patent text with occupational title text is used as a measure of identifying labor-augmenting innovations.
3. The data on occupational tasks describes the task content of occupations.[3] Since each occupation consists of numerous tasks and each task is described in one or two sentences, I group occupations by all the tasks they involve to get the full task data per occupation.

4.2 Methodology

The steps used to identify automation and augmentation innovations and the occupations they affect are outlined below and illustrated in Figure 1. Some steps are different based on whether documents embeddings are generated using GloVe [7] or sentence transformers.[4]

1. **Pre-processing**: For generating GloVe embeddings per document, this step strips punctuation, removes stop words, retain nouns and verbs, and lemmatizes each word. Pre-processing for sentence transformers involves only removing stop words and punctuation.

[1] One source of this data is this website: https://patentsview.org/.
[2] https://www.census.gov/topics/employment/industry-occupation/guidance/indexes.html
[3] https://www.onetcenter.org/dictionary/26.1/excel/task_statements.html
[4] The code for generating scores using these steps was not provided or taken from another source.

2. **Extract embeddings from words**: For generating GloVe embeddings, I use pre-trained word vectors from the Common Crawl (42B tokens, 1.9M vocab, uncased, 300d vectors, 1.75 GB download).
3. **Generating document vectors**: Calculate the TF-IDF weighted average of GloVe word vectors to generate a document vector. For BERT-base and MPNet-base sentence transformers, document vectors are directly calculated using the sentence-transformers package. The BERT-base sentence transformer model (bert-base-nli-mean-tokens) uses a BertModel as the transformer layer ('max_seq_length': 128) followed by a pooling layer. The MPNet-base sentence transformer model (all-mpnet-base-v2) uses the MPNetModel as the transformer layer ('max_seq_length': 384) followed by a pooling layer and a normalization layer.

   In this manner, a "document vector" is calculated for all CAI industry or occupation titles for each Census year in the sample and for all United States utility patents issued in the same time period.
4. **Measuring patent similarity to occupational tasks and titles**: Calculate cosine similarity for each patent-occupational title pair and each patent-occupational task pair using equation 1. To account for the fact that some types of patents have naturally low similarity scores (e.g. those using highly technical terminology such as chemical patents), the scores are normalized by subtracting the median score across occupations (or industries) for a given patent as shown in equation 2.

$$\text{cosine similarity} = S_c(P_i, O_j) = \frac{\vec{p_i}\vec{o_j}}{|\vec{p_i}||\vec{o_j}|} \tag{1}$$

$$\text{patent score}_i = \max(S_c(P_i, O_j)) - \text{median}(S_c(P_i, O_j)) \;\; \forall j \in J. \tag{2}$$

   The automation score of a patent is the degree to which a given patent automates work performed by human labor. To arrive at the automation score for a patent, I measure the cosine similarities between the document embedding of the patent summary text and the document embeddings of task descriptions for 760 occupations. The augmentation score measures the degree to which a given patent augments human labor. To arrive at this measure, I calculate the cosine similarities between the document embeddings of the patent summary text and occupational titles for 760 occupations. The approach for measuring the augmentation scores is similar to that for automation scores but the underlying occupational text contains titles of occupations instead of their task content as done in prior work [13].
5. **Identifying patent-occupation matches**: The top 5% most similar or highest adjusted textual similarity scores across patent × occupation pairs are retained as matches for patent p and occupation j. Lastly, the citation-weighted sum over patents issued in a given period is taken to obtain patent counts by occupation over time.

$$I_{p,j} = 1 \;\text{ if }\; S_c(P_i, O_j) > \sigma, \tag{3}$$

   where σ is the 95^{th} percentile of the similarity score distribution.

$$\text{occupational exposure}_j = arcsinh(\sum_p I_{p,j}) \tag{4}$$

4.3 Baseline

Prior work uses a Naive Bayes classification approach to assign patents into either automation or non-automation categories after manually labelling approximately 500 patents as automation/non-automation approach [12].[5] I use this classification as a baseline for identifying automation patents using different word and sentence embeddings.

5 Experiments

I compare patent-level scores obtained via different methods (GloVe embeddings, and embeddings from sentence transformer BERT-base and MPNet-base models) with the baseline and with each

[5]This data is provided by the authors: https://github.com/lpuettmann/automation-patents.

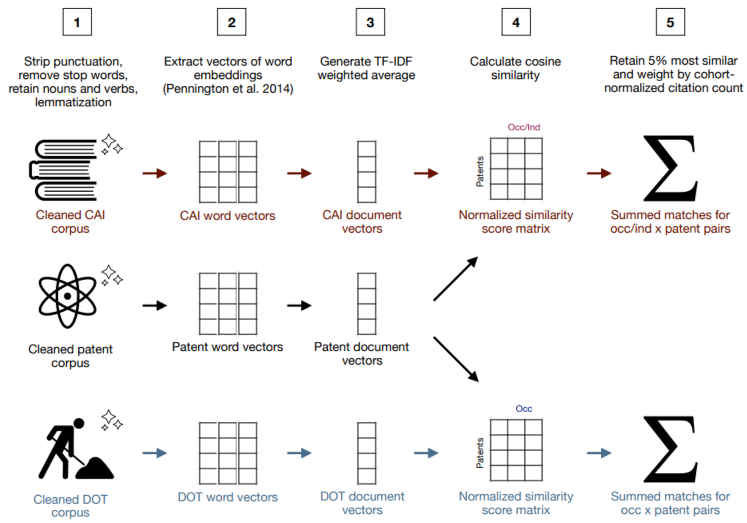

Figure 1: Generating automation and augmentation scores for patents. Source: [13].

other. To ensure whether the scores generated are capturing meaningful outcomes, I examine whether the occupation-level scores are predictive of occupational task content (i.e. the degree to which an occupation involves routine, abstract and manual work).

5.1 Data

To evaluate these measures, I analyze how they relate to the nature of jobs. To do so, I compile measures of the degree to which an occupation's tasks consist of routine, abstract and manual tasks based on prior work in this literature [14]. Prior work evaluates the tasks performed in various occupations based on the degree of routine and non-routine tasks involved in each occupation using the 1960 distribution of task input. The year 1960 is used as the base period for this standardization because it should primarily reflect the distribution of tasks prior to the computer era.

To link these task measures to my dataset, I rely on several crosswalks of US Census data. The United States Census records the detailed titles of workers' occupations. The publicly available Census data aggregates this occupation information and reports several hundred 3-digit occupation codes. The occupational classification system gets redefined for every decennial Census. In order to track detailed occupations over time, empirical work has to rely on crosswalks that match occupation codes from different Census years. I use the Occ1990dd Occupation System with 330 'occ1990dd' codes developed in prior work [15]. I then link the occ1990dd occupational codes to the 2010 Standard Occupational Classification (SOC) codes using a crosswalk file[6], and to the 2018 SOC codes (via the 2010 SOC to 2018 SOC crosswalk).

5.2 Evaluation method

This analysis is based on the assumption that automated machines are good at carrying out repetitive tasks and fail at complex abstract tasks as considered in prior work [14, 12]. Previous work shows that changes in routine-task intensity are predicted by investment in computer capital: The share of non-routine tasks increases, whereas that of routine tasks decreases following computer investment [14]. Prior work also shows that the relationship between automation patents and the routine-task index is positive: The larger the routine task share of an industry in 1960, the more automation technology was subsequently invented, patented and potentially used in that industry over the following decades [12]. Therefore, such an indicator appears to be capturing the same phenomenon as described by the literature on routine-biased technological change.

[6]Available at: https://www.ddorn.net/data.htm

To evaluate the occupation-level scores obtained via different pretrained embeddings, I examine how well these scores predict the degree to which an occupation's task consists of routine, abstract and manual work using linear regressions.

5.3 Results

Table 1 shows the automation scores obtained via different methods by industry. Note that the Naive Bayes approach is based on a binary classification into automation and non-automation patents whereas the other two approaches are based on the textual similarity of patent summary text to the text of occupational tasks.

Industry	Naive Bayes	GloVe	bert-base-nli-mean-tokens	mpnet-base
Chemical	0.16	0.87	0.72	0.52
Computers and Communications	0.98	0.92	0.73	0.55
Drugs and Medical	0.30	0.87	0.75	0.53
Electrical and Electronic	0.57	0.90	0.70	0.52
Mechanical	0.48	0.92	0.72	0.54
Others	0.37	0.92	0.73	0.56

Table 1: Mean (non-normalized) automation scores by industry for patents issued in 2010.

Table 1 shows that the GloVe and MPNet-base models follow patterns consistent with the Naive Bayes baseline and with each other (e.g. all three have the maximum score for the "Computers and Communications" industry). Previous experimental results show that MPNet achieves better results on similar tasks compared with previous state-of-the-art pre-trained methods (e.g., BERT, XLNet, RoBERTa) under the same model setting [9]. Moreover, GloVe-based scores have been shown in prior work to predict employment trends over time [13]. Given this, I further examine scores obtained via GloVe and MPNet. To check whether these scores change considerably from year-to-year, I calculated automation scores using the MPNet model for patents issued in 2011. After aggregating the patent level scores for both 2010 and 2011, I obtain the same scores by industry for the MPNet model (up to 2 decimal places) as mentioned in the last column of Table 1, which indicates that there is no yearly substantial variation in the scores obtained.

Table 2 shows the normalized patent-level scores (obtained via equation 2) by industry.

Industry	Naive Bayes	glove	mpnet-base
Chemical	0.16	0.082	0.327
Computers and Communications	0.98	0.070	0.309
Drugs and Medical	0.30	0.083	0.349
Electrical and Electronic	0.57	0.078	0.333
Mechanical	0.48	0.082	0.314
Others	0.37	0.083	0.321

Table 2: Mean normalized automation scores by industry.

Table 3 shows how the occupation-level automation and augmentation scores obtained via GloVe embeddings relate to the task content of occupations: routine, abstract and manual. As predicted, there is a statistically significant and positive relationship between occupations exposed to automation patents and their routine task content. This shows that more automation technologies were invented to automate work in occupations that consisted of a greater proportion of routine tasks. The statistically significant and negative relationship between occupations exposed to automation patents and their abstract task content along matches the prediction that automation technologies fail at complex abstract tasks. The linear regressions for augmentation scores are not statistically significant. One reason for this could be the lack of data: while mean number of preprocessed nouns and verbs derived from occupational task descriptions is 195.3 words, that for occupational titles is only 72.7 words.

Table 4 shows that scores based on MPNet-base sentence embeddings do not have similar statistically significant relationships as shown in Table 3. The MPNet-based model may be unable to capture these patterns because it uses all words in the document instead of only nouns and verbs as done during pre-processing for GloVe embeddings; consequently the MPNet-based model could be capturing greater

	Routine	Abstract	Manual
(Intercept)	3.76^{***}	3.18^{***}	1.06^{***}
	(0.25)	(0.22)	(0.15)
IHS_aut_patents	0.36^{**}	-0.38^{***}	0.24^{***}
	(0.12)	(0.11)	(0.07)
IHS_aug_patents	0.08	-0.13	-0.10
	(0.11)	(0.10)	(0.06)
R^2	0.10	0.15	0.10
Adj. R^2	0.08	0.13	0.08
Num. obs.	118	118	118
RMSE	1.99	1.77	1.17

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Relationship between Occupational Augmentation and Automation Exposure and Occupational Task Content based on normalized gloVe scores.

	Routine	Abstract	Manual
(Intercept)	4.34^{***}	2.24^{***}	1.14^{***}
	(0.23)	(0.19)	(0.13)
IHS_aut_patents	0.05	0.23^{*}	0.05
	(0.12)	(0.10)	(0.07)
IHS_aug_patents	0.23	-0.04	0.03
	(0.12)	(0.10)	(0.07)
R^2	0.03	0.04	0.01
Adj. R^2	0.02	0.02	-0.01
Num. obs.	136	136	136
RMSE	2.10	1.78	1.25

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Relationship between Occupational Augmentation and Automation Exposure and Occupational Task Content based on normalized mpnet scores.

noise while the GloVe embeddings are likely only capturing the similarities between meaningful words in each document. Another possible explainable could be that the MPNet model is based on a maximum sequence length of only 384 tokens whereas GloVe embeddings stem from all nouns and verbs mentioned in the summary text of each patent document. Collectively, Tables 3 and 4 show that automation scores based on GloVe embeddings better fit this prediction about routine and abstract tasks.

Model	Correlation
Naive Bayes	0.28
MPNet	0.05
GloVe	0.31

Table 5: Correlation of an occupation's routine task measure and associated automation patents.

Table 5 shows the correlation of the routine task measure of an occupation and the number of automation patents linked to that occupation. This further shows that the GloVe embedding method are better able to capture the routine task content of occupation relative to the MPNet-based model. It also suggest that the GloVe embedding method may be more predictive of routine tasks than the baseline Naive Bayes classification used in prior work [12]. However, the Naive Bayes correlation from prior work is based on industry-level measures (i.e. log of automation patents per industry and the average task content of occupations in that industry) whereas the GloVe and MPNet correlations are based on occupation-level measures.

6 Analysis

To further evaluate the patent-level scores obtained via GloVe embeddings, I plot the normalized automation and augmentation scores per patent, as shown in Figure 2. Figure 2 shows that automation and augmentation patents are positively related to one another.

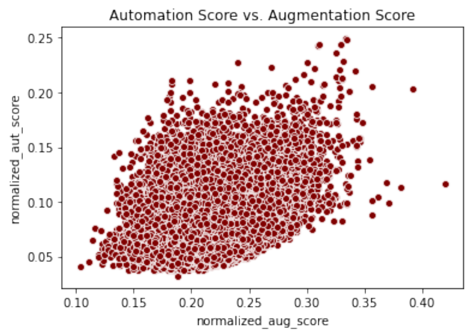

Figure 2: Relationship between patent-level normalized automation and augmentation scores.

I then examine the types of occupations linked to high automation patents and high augmentation patents. Interestingly, the types of occupations linked to high automation and high augmentation patents are quite different from one another, as shown in Figures 3 and 4, respectively. These figures are based on linkages between patents and occupations. However, each patent was only linked to the occupational task description or set of titles with which it had the highest similarity. Future work should link patents with all occupations and then select patent-occupation pairs that have a similarity score above a certain threshold to give a more comprehensive and accurate result of patent-occupation linkages since each patent may have applications (i.e. it may be automating or augmenting work) in more than one occupation.

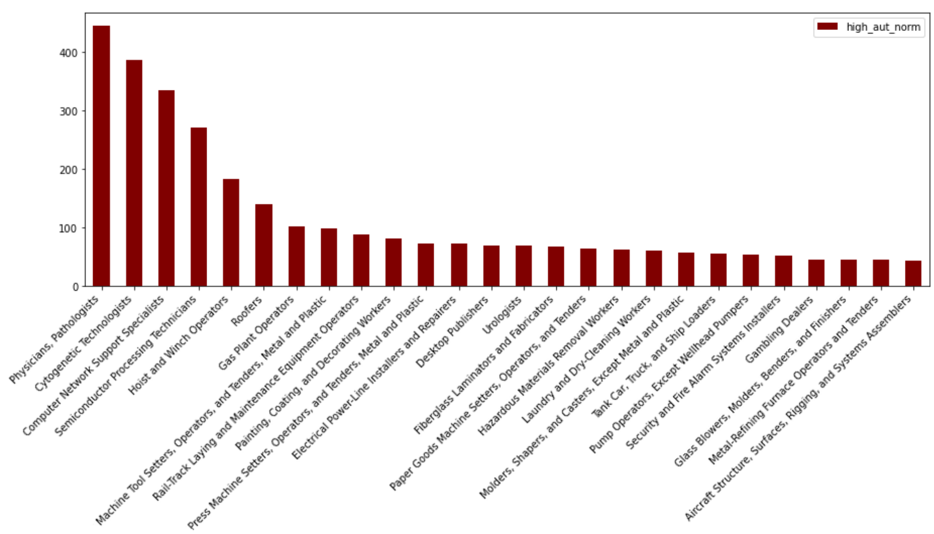

Figure 3: Occupational exposure to automation patents. The y-axis is the inverse hyperbolic sine of the number of automation patents.

This project also provides insight that can potentially explain why prior work finds different effects on labor market outcomes resulting from exposure to automation based on differences in the types of measures used. Prior work using the Naive Bayes classification method on classifying patents found that new automation technology per worker is significantly and positively related to employment gains in the same commuting zone [12], which paints a positive picture of the net employment effects of automation. On the other hand, recent work using Glove embeddings of patent documents finds that automation exposure predicts statistically significant declines in occupational employment.

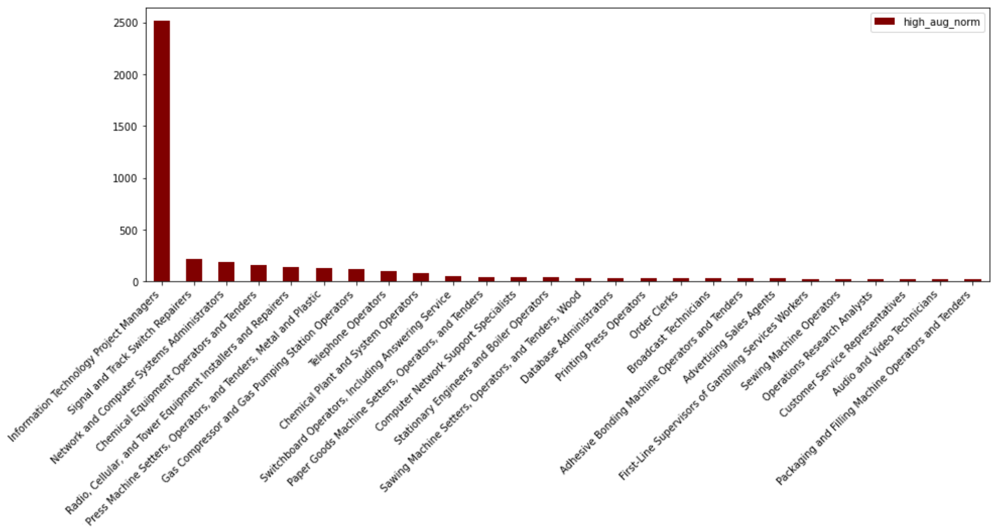

Figure 4: Occupational exposure to augmentation patents. The y-axis is the inverse hyperbolic sine of the number of augmentation patents.

In Table 6, I use a simple probit regression to see how the automation and augmentation scores obtained via Glove embedding methodology predict the likelihood of a patent being classified as an automat as per the Naive Bayes classification in prior work [12]. The results here show that being classified as an automat is negatively associated with automation scores and positively associated with augmentation scores, which suggests that the Naive Bayes classification is perhaps capturing augmentation technologies and therefore predicting a positive impact on employment growth. This comparison also suggests that the way we define our measures can significantly impact the types of labor market and other outcomes we find.

	automat
(Intercept)	-3.60^{***}
	(0.04)
normalized_aut_score	-26.11^{***}
	(0.25)
normalized_aug_score	28.13^{***}
	(0.21)
AIC	96725.23
BIC	96753.59
Log Likelihood	-48359.61
Deviance	96719.23
Num. obs.	94346

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Probit model predicting the likelihood of being classified as an automat [12] based on patent-level automation and augmentation scores derived from GloVe embeddings.

7 Conclusion

This project was aimed at identifying the types of innovations that automate and augment human labor using patent summary text data. From a dataset of 94,368 patents issued to US firms in 2010 and a comparison of both static and contextual word embedding techniques, the key finding based on this work is that automation scores obtained via GloVe embeddings-based document vectors are better able to predict the routine and abstract task content of occupations relative to contextual embeddings via MPNet-based sentence transformers. Additionally, while automation and augmentation scores for patents are positively related, the occupations linked to high automation and high augmentation patents are different from one another. Contextualizing these measures with prior work shows that the labor market impact of automation may be different depending on the definition of automation and the measures used. Future work that collectively links these measures to employment trends over time and compares the findings of measures obtained via different modeling techniques would be fruitful in better understanding how to evaluate these measures.

References

[1] Mitchell T. Brynjolfsson, E. and D. Rock. What can machines learn, and what does it mean for occupations and the economy? In *In AEA Papers and Proceedings, volume 108, pages 43–47*, 2018.

[2] M. Raj R. Seamans Felten, E. Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. In *Strategic Management Journal*, 2021.

[3] Michael Webb. The impact of artificial intelligence on the labor market. In *Working paper*, 2020.

[4] David Autor Joe Hazell Acemoglu, Daron and Pascual Restrepo. Ai and jobs: Evidence from online vacancies. In *NBER Working Paper 28257. DOI: 10.3386/w28257.*, 2020.

[5] Carl Benedikt Frey and Michael A. Osborne. The future of employment: How susceptible are jobs to computerisation? In *Technological Forecasting and Social Change, Volume 114, 2017, Pages 254-280, ISSN 0040-1625*, 2017.

[6] Percy Liang Mina Lee and Qian Yang. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *CHI*, 2022.

[7] Richard Socher Pennington, Jeffrey and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.

[8] Kenton Lee Jacob Devlin, Ming-Wei Chang and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT (1) 2019: 4171-4186*, 2019.

[9] Tao Qin Jianfeng Lu Kaitao Song, Xu Tan and Tie-Yan Liu. Mpnet: Masked and permuted pre-training for language understanding. In *34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.*, 2020.

[10] Morten Olsen Antoine Dechezlepretre, David Hemous and Carlo Zanella. Automating labor: Evidence from firm-level patent data. In *Working paper*, 2019.

[11] Dimitris Papanikolaou Lawrence Schmidt Kogan, Leonid and Bryan Seegmiller. Technological change and occupations over the long run. In *Working paper*, 2021.

[12] Katja Mann and Lukas Püttmann. Benign effects of automation: New evidence from patent texts. In *The Review of Economics and Statistics.*, 2021.

[13] Anna Salomons David Autor and Bryan Seegmiller. New frontiers: The origins and content of new work, 1940–2018. In *Working paper*, 2021.

[14] F. Levy Autor, D. H. and R. J. Murnane. The skill content of recent technological change: An empirical exploration. In *The Quarterly Journal of Economics 118(4), 1279–1333*, 2003.

[15] David Dorn. Essays on inequality, spatial interaction, and the demand for skills. In *Dissertation University of St. Gallen no. 3613, September. Data Appendix, pp. 121-138.*, 2009.