UNIVERSITY OF TECHNOLOGY SYDNEY

Changing Labour Market Dynamics in Australia: Skill Shortages, Job Transitions, and Artificial Intelligence Technology Adoption

by Nikolas Dawson

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Distinguished Professor Mary-Anne Williams,
Dr Marian-Andrei Rizoiu, and
Dr Benjamin Johnston

University of Technology Sydney
Faculty of Engineering & IT

June 2021

Certificate of original authorship

I, *Nikolas Dawson* declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science*, *Faculty of Engineering & IT* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

SIGNATURE: Production Note: Signature removed prior to publication.

[Nikolas Dawson]

DATE: 27^{th} June, 2021

PLACE: Sydney, Australia

Abstract

The Australian labour market is in the midst of significant structural changes. Emerging technologies, such as Artificial Intelligence (AI), are changing the demands for skills and tasks within jobs. Additionally, the economic crisis induced by COVID-19, in conjunction with other factors, have accelerated these adjustments. Three major issues for the Australian labour market are (1) skill shortages, (2) job transitions, and (3) AI adoption at the firm-level. This thesis by compilation addresses each of these issues in a series of four standalone papers. The first paper puts forward a range of indicators to detect skill shortages from a longitudinal dataset of online job advertisements (ads). The second paper develops a machine learning model that accurately predicts skill shortages from job ads data and employment statistics. The third paper conducts an in-depth case study of the journalism jobs crisis in Australia, examining both the changes in labour demand (using job ads) and labour supply (using employment statistics) from 2012-2020. Last, the fourth paper develops a novel method to measure the similarity between sets of skills from real-time job ads data. These similarity measures are then combined with other labour market variables to build a 'Job Transitions Recommender System' that accurately predicts transition pathways between occupations, validated against a longitudinal household survey. The same skills set similarity method is then used to construct a measure of new technology adoption in labour markets, showcasing AI.

DEDICATION

To my wife, Emma, for her ongoing love, patience, and support.

Acknowledgments

Undertaking doctoral research has been both the hardest and most rewarding endeavour of my career thus far. It has been a marathon to complete and a rollercoaster of emotions. While working towards my doctorate has developed skills and knowledge that I'll carry forward in my career, it has also taught me a significant lesson: *the most important thing to do is to think for yourself.* This is the foundation of great research or creating anything new of lasting value. A lesson that is simple to preach but difficult to practice.

I underestimated the amount of support required for completing a Ph.D. when I began my research. Thankfully, I have had plenty, for whom I would like to acknowledge here. First, to my principal supervisor, Mary-Anne Williams, who made my candidacy possible. Mary-Anne has enabled opportunities throughout my Ph.D. Most notably, arranging a research assistant position with the United Nations in Geneva for a project on the 'Economic Impacts of Artificial Intelligence'. I thank Mary-Anne for the opportunities and support she has provided – these opportunities have shaped me professionally and personally. I would also like to thank Benjamin Johnston, one of my co-supervisors, for some technical assistance and valuable feedback he provided early in my Ph.D.

To Marian-Andrei Rizoiu (Andrei), my other co-supervisor, I express my deepest gratitude. The research presented in this thesis and their corresponding publications would not have been possible without Andrei's guidance. Andrei is among the most intelligent and capable people I know. It has been a rare privilege to work with and learn from Andrei. I hope that we get the opportunity to work together in the future again.

Central to all of the research papers that I present in this thesis are job advertisements data, which have been generously provided by Burning Glass Technologies (BGT). Therefore, I would like to acknowledge and thank BGT for granting access to these data that enabled this research. Specifically, I would like to thank Bledi Taska and Davor Miskulin from BGT for establishing the data sharing agreement and for their valuable feedback on the various papers.

Last, I would like to thank my family and friends for being a constant source of support during this process. To my mother, Annmaree, who has created a foundation of support that has given me every chance to succeed. To my uncle, Berto, whose editing and feedback has been invaluable for the research papers presented in this thesis and other pieces of writing. And to my wife, Emma, whose love and support has been unwavering, even during the stresses of deadlines. Thank you for centring and helping me on this journey.

List of Publications & Statement of Contribution

Related to the Thesis (Published):

1. Dawson, N., MA. Rizoiu, B. Johnston, and MA. Williams. 2019. "Adaptively Selecting Occupations to Detect Skill Shortages from Online Job Ads." In *2019 IEEE International Conference on Big Data (Big Data)*, 1637–43 [120].

2. Dawson, N., MA. Rizoiu, B. Johnston, and MA. Williams. 2020. "Predicting Skill Shortages in Labor Markets: A Machine Learning Approach." In *The 4th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in BigData* (Co-located with 2020 IEEE International Conference on Big Data) [121].

3. Dawson, N., S. Molitorisz, MA. Rizoiu, and P. Fray. 2021. "Layoffs, Inequity and COVID-19: A Longitudinal Study of the Journalism Jobs Crisis in Australia from 2012 to 2020." *Journalism* [118].

Related to the Thesis (Under Review):

4. Dawson, N., MA. Rizoiu,and MA. Williams. 2021. "Skill-driven Recommendations for Job Transition Pathways." – submitted to *PLOS ONE* and awaiting second round of peer reviews as of the 25th of June 2021.

Not included in the Thesis (Published):

5. Gromov, A., A. Maslennikov, N. Dawson, K. Musial, and K. Kitto. 2020. "Curriculum profile: modelling the gaps between curriculum and the job market." In *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, pp. 610-614 [168].

6. Dawson, N., and MA Rizoiu. 2020. "Coronavirus Infecting Australian Jobs: Vacancy Rates down since Early February." The Conversation, March 22, 2020. http://theconversation.com/coronavirus-infecting-australian-jobs-vacancy-rates-down-since-early-february-134234.

Table 1: Papers and contributions by co-authors.

Title	Lead Author	Co-Author 1	Co-Author 2	Co-Author 3
Adaptively Selecting Occupations to Detect Skill Shortages from Online Job Ads.	Nikolas Dawson	**Dr. Marian-Andrei Rizoiu**, Lecturer at UTS and Ph.D. co-supervisor – assisted with the design of the research, data analysis, and manuscript editing.	**Dr. Benjamin Johnston**, Senior Lecturer at UTS and Ph.D. co-supervisor – provided feedback and technical assistance.	**Distinguished Professor Mary-Anne Williams**, UTS and Ph.D. supervisor – assisted with data collection and provided feedback.
Predicting Skill Shortages in Labor Markets: A Machine Learning Approach.	Nikolas Dawson	**Dr. Marian-Andrei Rizoiu**, Lecturer at UTS and Ph.D. co-supervisor – assisted with the design of the research, data analysis, and manuscript editing.	**Dr. Benjamin Johnston**, Senior Lecturer at UTS and Ph.D. co-supervisor – provided feedback and technical assistance.	**Distinguished Professor Mary-Anne Williams**, UTS and Ph.D. supervisor – assisted with data collection and provided feedback.
Layoffs, Inequity and COVID-19: A Longitudinal Study of the Journalism Jobs Crisis in Australia from 2012 to 2020.	Nikolas Dawson	**Dr. Sacha Molitorisz**, Lecturer at UTS – assisted with the design of the research, preparation of the manuscript, and journalism-specific expertise.	**Dr. Marian-Andrei Rizoiu**, Lecturer at UTS and Ph.D. co-supervisor – assisted with the design of the research, data analysis, and manuscript editing.	**Peter Fray**, Managing Editor at Private Media – assisted with the design of the research.
Skill-driven Recommendations for Job Transition Pathways.	Nikolas Dawson	**Dr. Marian-Andrei Rizoiu**, Lecturer at UTS and Ph.D. co-supervisor – assisted with the design of the research, data analysis, and manuscript editing.	**Distinguished Professor Mary-Anne Williams**, UTS and Ph.D. supervisor – assisted with data collection and provided feedback.	

THE ABOVE STATED INFORMATION IS CORRECT:

Production Note:
SIGNATURE: Signature removed prior to publication.

[Marian-Andrei Rizoiu]

Production Note:

SIGNATURE: Signature removed prior to publication.

[Mary-Anne Williams]

Production Note:

SIGNATURE: Signature removed prior to publication.

[Benjamin Johnston]

Production Note:

SIGNATURE: Signature removed prior to publication.

[Sacha Molitorisz]

Production Note:

SIGNATURE: Signature removed prior to publication.

[Peter Fray]

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Australian labour market has undergone seismic shifts since Federation [329]. Over the past two centuries, economic activity in Australia has moved away from agriculture and manufacturing, and towards service industries, with mining continuing to offer an important source of economic comparative advantage [219]. As the structure of Australia's economy has changed, so too has the share of employment across industries and the nature of work itself. The proportion of Australian workers employed in service-based industries has grown from approximately 52% in 1910 to 84% in 2010 [111]. These structural changes are driven by a range of factors, such greater disposable income and changing labour demographics that increases the demand for services [207], the industrialisation of Asia, increased trade activities, and economic reforms [111, 218, 282]. Technological advances have also played a pivotal role in shaping modern labour markets [68, 84], including Australia [182]. The introduction of new tools and technologies have altered the nature of work by improving labour productivity, thus changing the demand for skills and tasks within jobs [80, 87, 194]. While naming conventions of standardised occupations remain relatively unchanged over the past five decades [25], the underlying skill demands within these occupations have evolved significantly [50, 182, 342].

Central to the works presented in this thesis is the concept that jobs (or standardised occupations) consist of labour tasks that require skills, knowledge, and abilities [7, 47, 242]. Here, 'skills' refer to the proficiencies developed through training and/or experience [263]; 'knowledge' is the theoretical and/or practical understanding of an area;

'abilities' are the competencies to achieve a task [157]; and 'occupations' are standardised jobs that are an amalgamation of skills, knowledge, and abilities that enable individuals to perform a set of tasks that are required by their unique occupation. This research area is referred to as the 'Task-Approach', which was introduced by Autor, Levy and Murnane [47] and further developed by Acemoglu and Autor [7], and Autor and Handel [44]. Essentially, jobs (or standardised occupations) can be defined by their constituent task demands, which are specific work activities that require skills, knowledge, and abilities. Further, the 'Task-Approach' shows that technologies do not substitute or augment labour at the job level, rather it is tasks within jobs that are substituted or augmented by technologies. Therefore, the granular elements of jobs contain rich information about the changing nature of work within occupations and labour markets. By analysing these micro components of jobs, we can conduct in-depth analyses of how jobs are evolving, and develop a greater understanding of labour market trends.

This thesis examines some of the major issues related to the changing labour market dynamics in Australia from 2012 to 2020, drawing from granular skills data in concert with other occupational data. The main research focus areas are detecting and predicting **skill shortages**, analysing occupations in the midst of significant disruptions and developing systems to assist with **job transitions**, and measuring **Artificial Intelligence technology adoption** within the Australian labour market. The research presented in this thesis draws upon a corpus of 8,002,780 online job advertisements (ads) in Australia from 2012-01-01 to 2020-04-30, courtesy of Burning Glass Technologies (BGT). The dataset is organised into a structured format, where key features have been extracted and standardised from each job ad text description. These include features such as occupational classifications, standardised industries, salary ranges, education requirements, experience demands, and skills (among others). In this dataset, 'skills' also encompass 'knowledge', 'abilities', 'tools and technologies', and some 'tasks'. This longitudinal dataset provides unique insights into the evolving labour demands of Australia. The job ads data are used for analysis in all four research chapters (or standalone papers under the central thesis theme).

While job ads provide a useful proxy for labour demand [101, 120, 227], they reflect only the demand-side of job markets and do not reflect labour supply. Just because a job is advertised and employers signal specific labour demands does not mean that the vacancy is filled. This is an important consideration for many problems in labour economics. For instance, skill shortages occur when the labour demand for specific skills exceed the supply of workers who possess those skills at a prevailing market wage [180, 196].

Therefore, it is critical to combine labour demand and labour supply data when analysing skill shortages, so that disequilibria between supply and demand of skills can be detected. Similarly, job transitions reflect actual movements of individuals between jobs, which is a supply-side measure of labour mobility. Labour demand data, however, can complement these data sources and provide explanatory features that represent job transitions, as will be shown in Chapter 6. I have combined labour market data from both demand and supply sources for three of the chapters in this thesis to strengthen the analyses.

1.1 Identified research gaps

After completing a comprehensive review of the literature in Chapter 2, I have identified a number of research gaps regarding Australia's changing labour market dynamics. Labour economics as a discipline has traditionally relied upon theoretical models that are specified with prescriptive assumptions to describe reality. While these modelling approaches can be useful, they can also be overly simplistic, neglecting the complexities of economic systems. Data-driven modelling approaches offer new research opportunities for established and growing problems in labour economics. The combination of greater access to economic data proliferated by the Internet, improvements in modelling techniques, such as machine learning, and increases in computational power have opened a plethora of data-driven research opportunities. These include opportunities and gaps in the areas of skill shortages, job transitions, and AI technology adoption.

For skill shortages, I have identified that there are research gaps to implement data-driven methods to *detect*, *predict*, and *understand* occupational skill shortages. The application of data science and machine learning methods offers data-driven approaches to analyse these established research problems. These methods allow for the productive use and analysis of the large labour market datasets currently available, without overly specifying the models with prescribed assumptions.

Similarly, there are clear research opportunities to examine job transitions with data-driven methods. By combining labour demand and labour supply data sources, data science techniques can be applied to comprehensively *assess* the disruptions within occupations and to *identify optimal transition pathways* for moving between occupations. In particular, the skills data extracted from job ads provides a rich data source to analyse job transitions according to their underlying skill demands and the transferability of these skills. This research gap represents a significant opportunity to develop systems to assist workers faced with the often distressing challenges of transitioning between jobs.

The final research gap identified for this thesis is using job ads data as *leading indicators of AI adoption*. As discussed in Chapter 2, there have been many significant studies on the potential impacts of AI on labour markets, including the Australian labour market [136]. However, most of these studies assume a growing rate of AI technology adoption by firms that underpin their forecasts. While it is likely that AI technology adoption will continue to accelerate, adoption and diffusion of emerging technologies is not automatic [68]. Having reviewed the literature, it was apparent that measuring AI adoption was an under-served area and also an essential factor that will affect the pace and extent of the impacts that AI will have on labour markets. While the adoption of emerging technologies are influenced by a range of factors, as discussed in Chapter 2, access to skilled labour for firms to make productive use of the emerging technologies is essential [68, 93]. Therefore, analysing job ads provides a rich source of information on the demand for AI-related skills and occupations, which is a leading indicator of firms acquiring the skilled labour to adopt and implement AI technologies.

These identified research gaps underpin the aims and subsequent contributions of this thesis.

1.2 Aims and contributions of this research

The aim and central theme of this thesis is to analyse the changing labour market dynamics in Australia from 2012 up to 2020. More specifically, the focus is on three major issues pertaining to the Australian labour market: (1) skill shortages, (2) job transitions, and (3) the demand for skills that enable AI adoption at the firm-level.

In the research related to skill shortages, these works focus on three open problems. First, methods and metrics to *detect* occupational skill shortages from real-time job ads data. Second, the ability to accurately *predict* occupational skill shortages in advance. And third, using a data-driven approach to *understand* which labour market features are most predictive of occupational skill shortages.

The objectives for the research concerning job transitions are twofold. First, to demonstrate methods and metrics for comprehensively analysing an occupation known to be currently experiencing structural disruptions and facing forced job transitions. Second, to develop a system that assists workers to efficiently transition between jobs according to their underlying skills.

Last, there are two aims for the AI adoption research. The first aim is to conduct a longitudinal analysis of the labour demands for Data Science and Analytics (DSA) occupations in Australia. DSA occupations are considered the occupational group most responsible for implementing and making productive use of AI technologies at the firm-level [70, 227]. Therefore, understanding the demand growth of DSA occupations can provide leading insights into the pace by which AI technologies are being adopted and diffused. The second aim is to develop a leading indicator of AI adoption using granular skills data. Here, the objective is to build an indicator that prioritises for skill importance and dynamically adjusts for changing skill demands over time.

The research presented in this thesis makes contributions to each of these aims. Chapter 3 puts forward a series of methods and metrics to detect skill shortages from job ads data. Additionally, this work applies these methods and metrics to assess the extent of skill shortages for DSA occupations in Australia, which have previously been shown to be in shortage [123]. Not only does this occupational case-study validate the usefulness of these methods and metrics for detecting skill shortages, it also provides a longitudinal account of the changing labour demands for DSA occupations. This contributes important insights into the increasing demand for DSA skills and jobs, and offers leading insights on the primary occupational group responsible for implementing AI technologies in firms. Chapter 4 implements a high-performing machine learning approach to predict occupational skill shortages one-year in advance, using both labour demand and labour supply data sources. In addition, feature importance analysis is conducted to understand which variables are most predictive of occupational skill shortages and a metric is developed to determine the most important skills within an occupation. Chapter 5 provides a longitudinal analysis of the journalist occupation in Australia. Journalists have been selected for this in-depth case-study due to the immense structural changes experienced in this occupation from 2012 to 2020. This work uses data science and machine learning methods to provide a thorough account of the changing labour dynamics of Australian journalists. It also identifies likely transition pathways to other occupations according to underlying journalism skills. Chapter 6, the final research Chapter, contributes a novel methodology to calculate the similarity between occupations using their skill sets from job ads data. The outputs from the method are then combined with other labour market variables to build a recommender system for identifying the optimal transition pathways between occupations. Not only are the results accurate (Accuracy = 76%) but the system also accounts for the asymmetries between job transitions (it is harder to transition in one direction than the other). The skill set similarity method is then further applied to develop a leading indicator for new technology adoption (using AI as the example). Collectively, these four research papers (presented here as Chapters) provide novel and meaningful contributions to the specific areas of skill shortages, job transitions, and AI adoption, and the broader literature on the changing labour market dynamics in Australia.

1.3 Thesis structure: A Ph.D. by compilation of papers

This dissertation is presented as a 'thesis by compilation' of research papers. A thesis by compilation is a single manuscript that consists of distinct works that are either published or publishable under a central research theme [325]. During my Ph.D. studies, I have successfully published three papers at highly respected conferences and journals, I currently have another paper under-review at a high-impact journal, and I was selected as an 'OECD Future of Work Research Fellow' for my doctoral research on job transitions and AI adoption. Therefore, given my focus on publishing research throughout my Ph.D. studies, a 'thesis by compilation' was an ideal structure for presenting my doctoral research. As the university does not have a prescribed structure for a 'thesis by compilation' [325], I present my four research papers as independent Chapters, which all relate to the central theme of *'The changing labour market dynamics in Australia from 2012-2020'*. The Materials and Methods are presented in each of the Chapters and there are general Literature Review and Conclusion Chapters. This dissertation is organised into the following seven Chapters:

- Chapter 2 is a Literature Review of Australia's labour market dynamics, the changing nature of work and skills, job transitions and skill mismatches, and AI technologies and their potential impacts on jobs. Each of the four research papers also include a review of the relevant literature. Therefore, to avoid repetitive content, the literature review presented in Chapter 2 relates more broadly to the central theme of this dissertation, whereas the literature reviews in each of the paper Chapters are more targeted to the research objectives outlined in the papers.

- Chapter 3 presents the paper entitled 'Adaptively selecting occupations to detect skill shortages from online job ads', which was published in the *2019 Proceedings of the IEEE International Conference on Big Data* [120].

- Chapter 4 showcases the paper entitled 'Predicting Skill Shortages in Labor Markets: A Machine Learning Approach'. This research has been published in the

Proceedings of The 4th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in Big Data as part of the *2020 Proceedings of the IEEE International Conference on Big Data*.

- Chapter 5 presents ‘Layoffs, Inequity and COVID-19: A Longitudinal Study of the Journalism Jobs Crisis in Australia from 2012 to 2020’, which has been published in the *Journalism* journal – a highly regarded International journal and among the leading journals in the Journalism discipline.
- Chapter 6 consists of the research paper entitled ‘Skill-driven Recommendations for Job Transition Pathways’. The paper is currently under-review at the *PLOS ONE* journal.
- Chapter 7 discusses future work, limitations of this research, and final conclusions. As each of the research Chapters consist of their own in-depth Discussion sections, this Chapter focuses on lines of inquiry that could advance the research presented in this thesis and related works that have recently emerged.

Chapter 2

Literature Review

2.1 Australia's changing labour market

The literature reviewed in this Chapter relates to the central theme of the thesis on the changing labour market dynamics in Australia. This begins with an overview of Australia's labour market, the changing nature of work, and how displaced workers are supported in Australia. Then there is a review of the key literature related to job transitions and skill mismatches that affect labour mobility. Next, is a non-technical overview of AI technologies and its potential to become the next significant group of 'General Purpose Technologies'. This is followed by an overview of the literature on the potential impacts of AI on labour markets. And last, is a review of the factors that affect AI adoption by firms.

As the Chapters 3-6 are all standalone research papers, each of these Chapters also contain sections on related works. However, to avoid repetition, the reviews of related works of Chapters 3-6 have not been duplicated in this Chapter. Instead, this Chapter provides a broad review of the literature related to the central theme of this thesis and the research Chapters provide more targeted reviews concerning their specific research focus.

2.1.1 Australia's labour market performance

Overall, Australia is considered an advanced economy that has enjoyed relatively stable economic growth and a robust labour market [261]. Prior to the COVID-19 pandemic, Australia had experienced almost 29 years of uninterrupted economic growth, which is the longest on record among the OECD advanced economies [316]. Similarly, the **headline unemployment** rate remained consistently low during the 2000's, before jumping above 7% as a result of the COVID-19 crisis [33], which was still slightly lower than the OECD average [266]. The blue line of Fig. 2.1 (a) shows the monthly unemployment rate in Australia [33], with clear recent spikes during the Global Financial Crisis (2008-2009) and the COVID-19 pandemic (2020).

Beneath the averages and headline statistics, however, are signs that Australia's labour market is undergoing structural changes. Fig. 2.1 (a) shows the diverging trends of unemployment and underemployment rates in Australia. Specifically, that while the rate of unemployment has trended down since 1993, the underemployment rate (orange line in Fig. 2.1 (a) has consistently trended upward since 1978, and sharply increased following the onset of the COVID-19 pandemic.

Youth unemployment and underemployment rates have stubbornly persisted above the headline rates. Fig. 2.1 (b) shows that youth unemployment (light blue line) exceeded 16% and youth underemployment (light orange line) exceeded 23% in 2020 following the COVID-19 outbreak [33]. Dhillon and Cassidy identify a number of factors contributing to these persistently higher unemployment and underemployment rates in Australia [132]. First, as younger workers are more likely to be new entrants to the labour market, their skills and work experiences are typically less developed. This exposes young people to cyclical swings in the economy and structural changes in the labour market. Second, increased rates of educational attainment and greater numbers of younger migrants have caused structural changes the labour supply of younger workers in Australia. Participation of people aged 15-24 in full-time education has steadily increased since 1988 and more than half of the net migrants in the past decade have been less than 25 years of age [132]. Last, labour demand in Australia has experienced a gradual shift towards part-time and more flexible work arrangements (as seen below in Fig. 2.2 (a) with the increasing share of part-time workers for males and females). Even before the labour crisis induced by COVID-19, the proportion of full-time equivalent employees to the total employment population had steadily declined from 84% in 1979 to 68% in 2018 [33]. This 'casualisation' of labour demand in Australia (colloquially referred to as the 'gig economy') has contributed to the rise in the youth underemployment

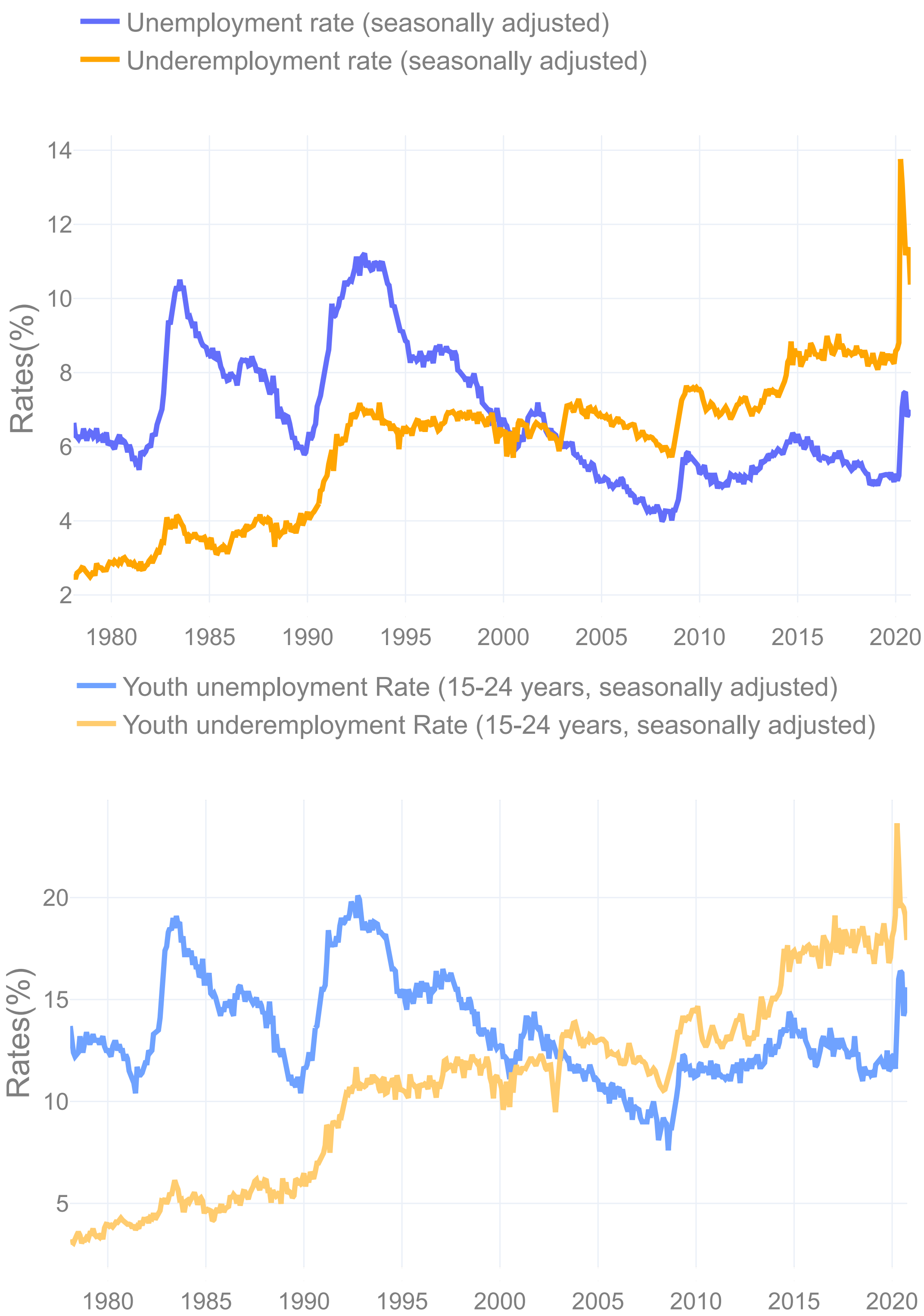


Figure 2.1: Monthly unemployment and underemployment rates in Australia from 1978 to 2020 for the (a) total Australian workforce and (b) youth workforce aged 15-24 years from the ABS [33].

rate [132].

The **gender** make-up of Australia's labour market has also undergone major shifts. Female participation in the labour market has increased significantly since 1978 [33]. Fig. 2.2 (a) shows the growth of females employed in Australia for both full-time (dark pink) and part-time (light pink) workers, which is visibly greater than the growth of male workers over this period. However, the majority of the female employment growth has been attributed to part-time employment, with average underemployment rates for females in Australia persisting at higher levels compared to men [33, 39]. In relative terms, Fig. 2.2 (b) highlights the high levels of growth of the employment-to-population ratio for females (pink line) in Australia aged 15-64 years, as opposed to the declining male employment-population-ratio (blue line). While female labour participation and educational attainment improved from 1978-2020, many inequalities persist for women working in Australia [258]. Gender wage gaps continued across of all Australian States and Territories from 1983-2013 [201] and, as of May 2020, female wages were, on average, 14% below male wage levels [339]. Also, compared to other developed economies, Australia has had slightly lower than average participation in formal childcare for children aged 0-2 years, which has contributed to the rising and relatively high levels of part-time employment for Australian women [258].

2.1.2 The changing nature of employment and skills in Australia's workforce

Australia is experiencing a pronounced structural shift away from a goods-producing economy to a service-based economy [181]. A major reason for this shift is that Australian households have experienced significant growth in their real incomes, which have more than doubled since the 1960s. As incomes rise, households are more likely to consume greater levels of services (such as education, health, and hospitality) than goods [207].

As seen in Fig. 2.3, service sectors have experienced the largest levels of employment growth in the Australian economy over the past two decades [127]. This has seen the Australian economy move away from the primary industries of Agriculture, Manufacturing, and Mining, and towards more service intensive sectors, such as Healthcare, Professional Services, and Education. These employment shifts towards a service intensive economy are consistent with the trends experienced by other advanced economies, as discussed in Section 2.3.7 of this literature review.

There are multiple factors contributing to this orientation towards a service-based

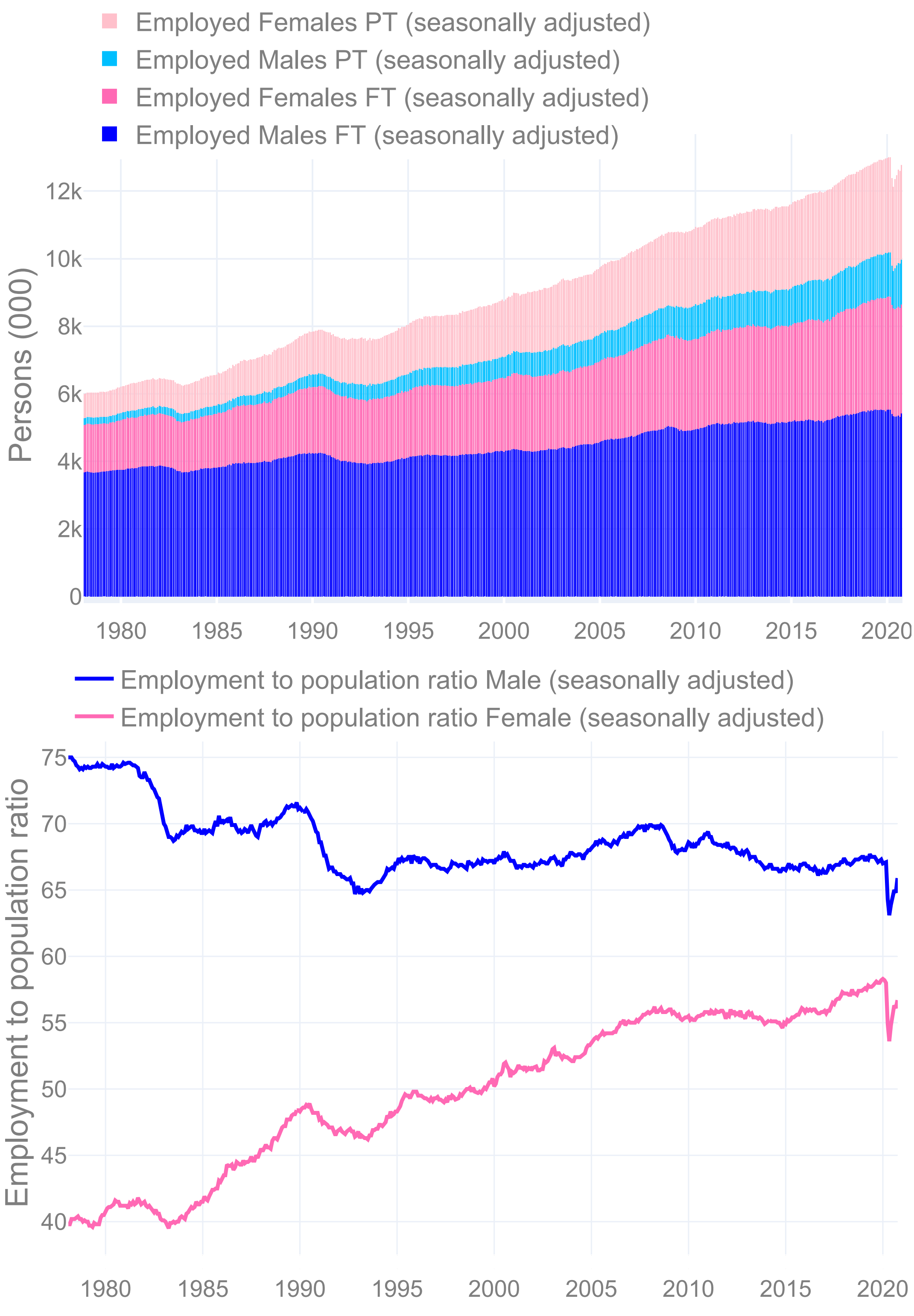


Figure 2.2: (a) Full-time (FT) and part-time (PT) monthly employment levels (000's) by gender in Australia (b) monthly employment to population ratio by gender in Australia.

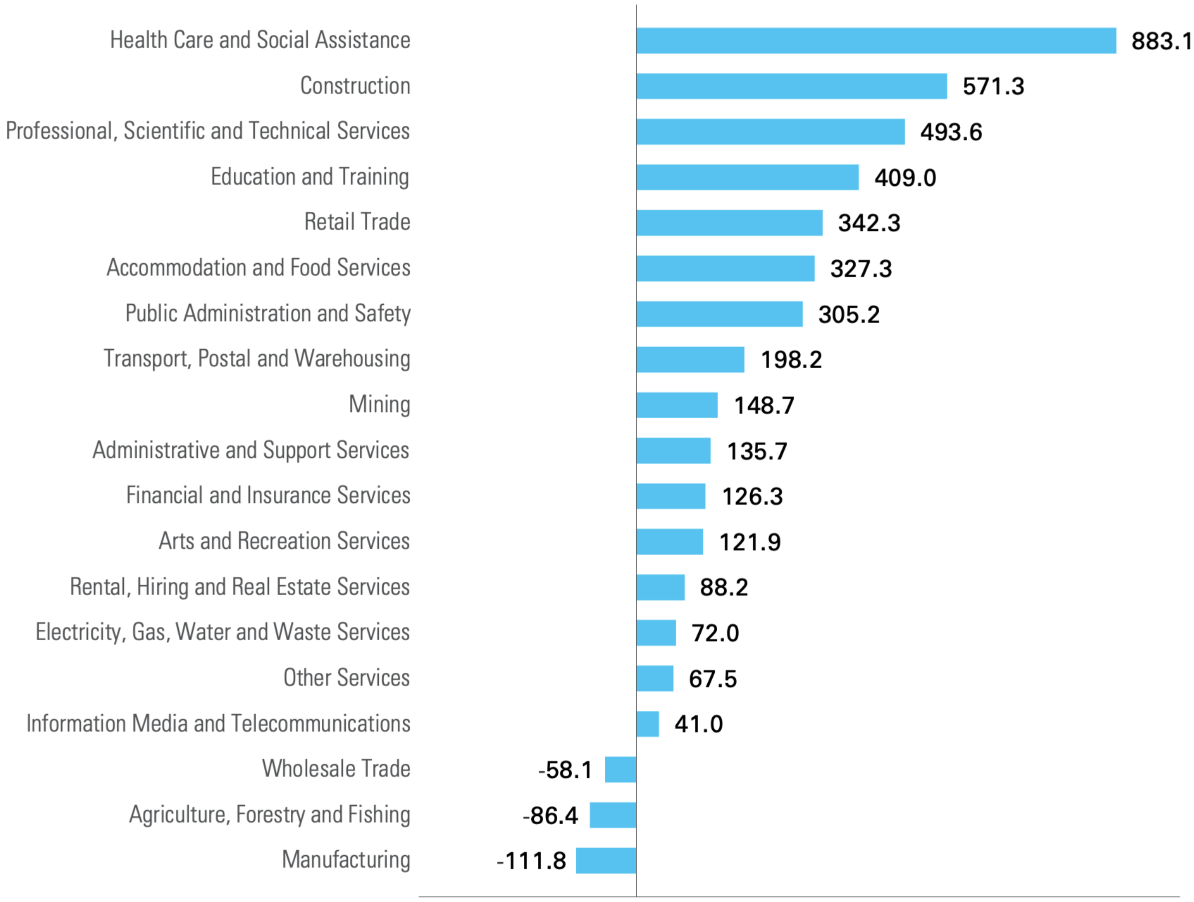


Figure 2.3: Employment change by industry over 20 years from to May 2018 (000's).

economy and labour market in Australia. First, Australia's ageing population places greater demand on aged care and health services, further orienting the labour market to service intensive economy [200]. Second, improved technological capabilities have enabled the goods-producing industries to increasingly automate production processes and outsource non-core activities to the business services sector (for example, financial services) [47]. This has increased specialisation within the economy, thus creating demand and new forms of services [181]. Third, as per capita income has increased in Australia, so too has the propensity for consumer preferences to increasingly orient towards services over goods [207].

These structural economic changes shift the skills demanded in the labour market. As seen in Fig. 2.4 (a) from analysis performed by the Reserve Bank of Australia (RBA), non-routine and cognitive skills as a percentage of total employment has increased in Australia from the 1980's until 2019 [182]. Both cognitive and manual non-routine jobs have experienced positive growth as a percent of total employed, whereas routine jobs

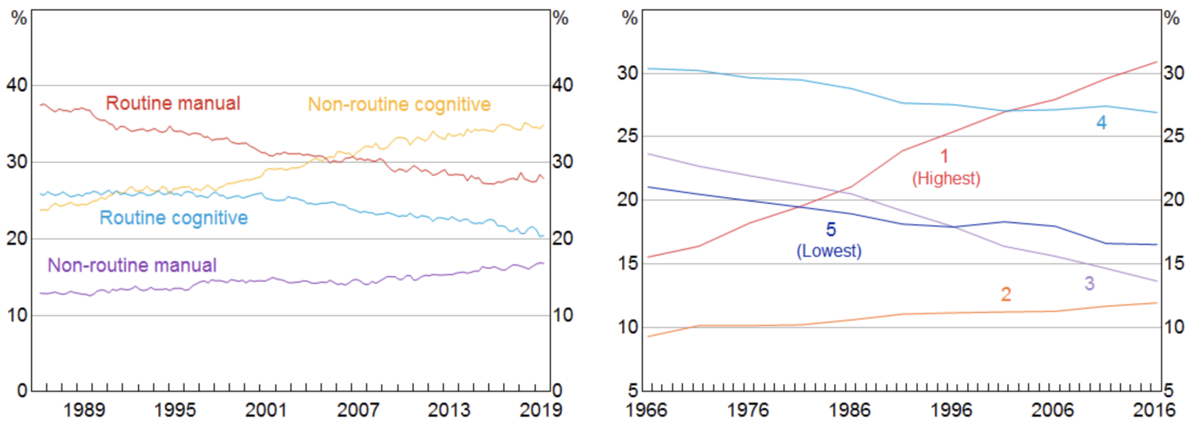


Figure 2.4: Australian (a) employment by skill type as a percentage of total employment and (b) employment by skill level as a percentage of total employment.

have gradually declined. This trend is consistent with other advanced economies [92, 151, 155]. Similarly, further RBA analysis shows in Fig. 2.4 (b) that employment for occupations that require the highest levels of skills (red line - number 1) had doubled from 15% in the 1960's to above 30% in 2016 [181]. Whereas the shares of employment for the lowest three skill levels (numbers 3-5) have gradually declined over this same period. One potential explanation for this phenomenon is that lower-skilled labour tasks within jobs are more susceptible to automation [47, 87, 136, 151] (this 'Skill-Biased Technological Change' theory is discussed in detail in Section 2.2 and Section 2.3.7). Occupations with lower skill requirements tend to consist of routine tasks that can be more readily codified by machines [7]. Whereas higher skilled occupations tend to require more non-routine labour tasks that are not as easily automated. Further, technologies that automate labour tasks generally enable higher skilled workers to become more productive and pursue more non-routine tasks [110]. Supporting this argument is the increasing prominence of service sector industries in advanced economies, such as Australia [182]. Occupations within service sector industries (such as 'Health' and 'Professional, Scientific and Technical Services') tend to require higher levels of non-routine and socio-cognitive labour tasks, which are difficult to automate [47].

This growing demand for high skilled labour is reflected in labour market outcomes. Fig. 2.5 shows that Australian workers with higher educational attainment (and presumably higher levels of skill) experience lower unemployment rates and higher participation rates [127]. Workers with higher educational qualifications also earn a wage premium, although this premium has lessened over the past two decreased [250].

The increasing demand for higher-skilled labour is having somewhat of a polarising

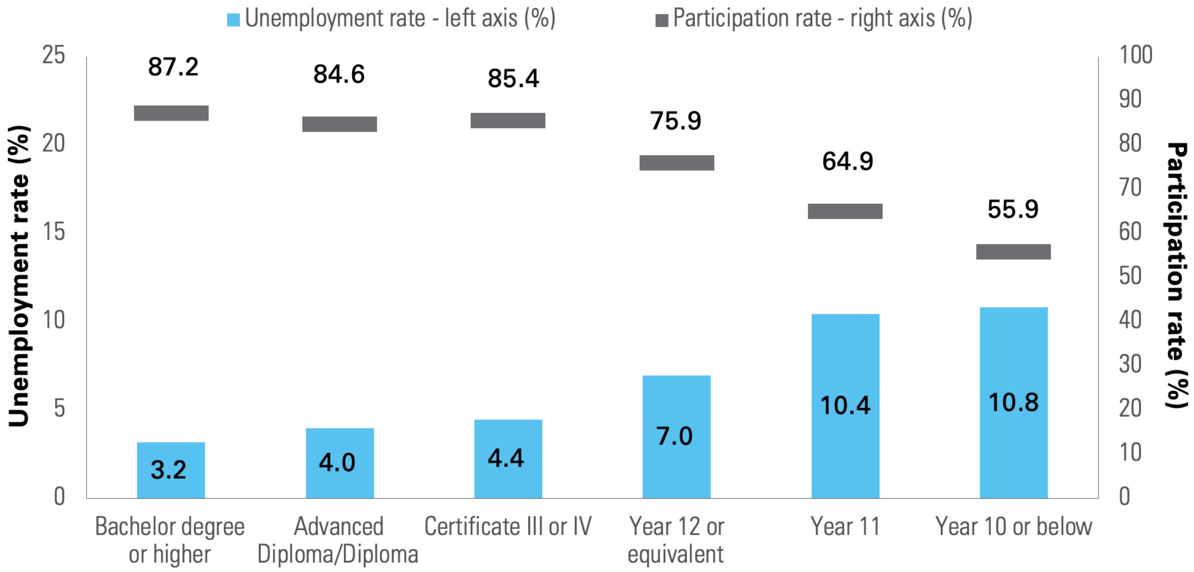


Figure 2.5: Australian unemployment and labour market participation in 2017 by educational attainment.

effect on the labour market. Workers that successfully upskill or reskill are able to join the ranks of higher skilled labour and meet this demand. Workers that are unable to make this transition, however, run the risk of either joining the pools of lower skilled labour or disengage from the labour market [87]. This claim is reflected in the increasing competition for lower skilled jobs. Based on a 2017 survey conducted by the Department of Jobs and Small Business [127], employers received an average of 19 applicants per lower skilled job, but only an average of 3 applicants were awarded an interview. Just over half employers (51%) considered experience essential for lower skilled vacancies, yet the majority of lower skilled job seekers lacked recent work experience [127]. A combination of increased supply and decreased demand is resulting in greater competition for lower skilled work.

The structural economic and labour market shifts discussed above have important implications for the future of work and economic outcomes in Australia. Technological changes are one of the principal factors influencing these structural shifts. And AI represents a group of technologies that could instigate such widespread technological changes (see Section 2.3.5 below for arguments and evidence). Therefore, how well the Australian economy manages these structural adjustments will depend significantly on how smoothly labour markets transition to fulfil new labour demands.

To assist with the efficient transitioning displaced labour from old industries to the new, governments utilise Labour Market Programmes that are designed to speed up this process. The following subsection provides a review of Labour Market Programmes and how they operate in Australia. It will also define displaced workers as a subset of the unemployed population, and the interaction of displaced workers with Australian Labour Market Programmes.

2.1.3 Labour Market Programmes: supporting displaced workers to transition

Dynamic labour markets are an important part of the 'creative destruction' process. Labour market dynamism efficiently reallocates labour from shrinking firms to growing firms, which helps to lift economic growth and standards of living [210]. However, dynamic labour markets require robust economic conditions to enable this quick and fluid transition of labour to meet growing demands [261]. These conditions depend upon a range of aggregate labour market variables, such as skill and educational levels, demographics, wages, aggregate demand, and several more [328]. If these conditions are not fulfilled, the process of labour reallocation can be difficult and inefficient, resulting in adverse labour market outcomes. When labour transition rates are slow, individuals can experience hardships, such as significant losses in wage income from unemployment or underemployment [261]; similarly, macroeconomic performance can be impeded by failing to make productive use of labour and the associated costs borne to the State.

Differentiating displaced workers from the unemployed population. Labour displacement is broadly defined as a process of involuntary job separation caused by structural changes to economic conditions [3, 75]. Thus, the causes of job separation are initiated by the employer in response to deteriorating economic conditions and are not caused by a worker's individual performance. Fallick further elaborates three prerequisites for job displacement [141]:

1. There is a structural change to economic conditions that causes the job separation (for example, technological change) rather than a cyclical downturn or the performance of an individual firm;

2. The prospects for returning to comparable jobs in the same industry with an equivalent wage are low; and

3. The workers were established in the industry where they were formerly employed, which is typically measured by a tenure of one or more years.

The Australian Bureau of Statistics (ABS) refers to displaced workers as 'retrenched' (referred to as displaced in this thesis). However, until the reporting period of August of 2014, the ABS did not differentiate displaced workers between those of whom were fired for under-performance and those of whom lost their job due to adverse economic conditions (hence, displaced). This caused challenges in isolating displaced workers in labour market statistics [241]. As a result, the review of Australian Labour Force statistics in 2012 [24] identified the importance of frequently measuring labour displacement, which excludes persons dismissed from their job. According to the ABS [26], the new definition is measured by:

> *"...the total number of persons who ceased a job during the last three months because they were either:*
>
> - *Retrenched, made redundant, employer went out of business, no work was available; or*
> - *Self-employed persons whose business closed down for economic reasons, including went broke, liquidated, no work, no supply or demand."*

Labour market programmes for displaced workers. Labour market programmes (or policies) are government initiatives designed to assist people that are unemployed to become re-employed as quickly as possible. Different approaches are distinguished as being either passive or active:

- *Passive* labour market programmes are less targeted and interventionist, such as income support benefits, pension payments, and concessions [254]. Passive measures can be unconditional or conditional, based on a particular eligibility criteria (for example, time limited or income means tested).
- *Active* labour market programmes are more targeted interventions, such as Job search and self-help assistance; Training programmes; Job placement services; Intensive support for disadvantaged job seekers; Job creation strategies, such as Intermediate Labour Market programmes that use social enterprises or work experience programmes; and Job subsidies to stimulate labour demand from employers.

The majority of advanced economies implement a combination of passive and active labour market programmes [261]. Passive measures, such as unemployment benefits, provide subsistence compensation to ensure income security. Whereas active strategies aim to accelerate the re-employment process through direct intervention. While both approaches have shown to have varying degrees of success for general unemployed populations [96], rigorous evaluation evidence on what works for displaced workers remains limited [261]. This is partly because many national labour market programmes do not adequately differentiate displaced workers from the unemployed population (such as Australia). It is also because displaced workers are often a heterogeneous and diverse group (for example, the distribution of education, skill, and socioeconomic background of displaced workers can be very broad). Both these characteristics add the complexities of data collection and, therefore, which support mechanisms have the highest re-employment outcomes for specific groups of displaced workers.

Nonetheless, there are some evaluation studies that measure the efficacy of outcomes for displaced workers. Barnow and Smith provide one of the largest studies targeting 'dislocated' workers (equivalent to displaced) from the United States (US) [51]. The study tracks trade-displaced workers and the active labour market programmes funded by the US Department of Labor and the Trade Adjustment Assistance program. There is little evidence to suggest that these targeted employment and training interventions were cost-effective measures for re-employing displaced workers [51]. However, as the authors note, there are significant flaws in the available data and therefore reasons to be sceptical of any general conclusions. Firstly, many of the displaced workers in the study had been unemployed for extended periods, whereas early intervention is more likely to be effective in achieving re-employment outcomes. And secondly, due to the fragmented nature of US labour market programmes, many of the 'unserved' displaced workers in the control group actually received similar services through other means.

The OECD Back to Work research is a series of in-depth country reviews of labour market programmes that target the re-employment of displaced workers from nine advanced economies [254]. While the strength of evidence varied between countries (for example, some countries differentiate displaced workers more clearly than others in national statistics), the country reviews are highly instructive. For instance, the labour market programmes offered to displaced workers in Canada (a country with similar demographics and socioeconomic status to that of Australia) provided compelling insights [252]. Specifically, the review found evidence that job-search assistance, conditional wage subsidies, and targeted training all increased post-displacement outcomes for the participation, employment, and earnings of displaced workers. Additionally, early intervention was shown to yield greater post-programme gains in both employment and earnings. While these reviews provide insight, one cannot discount the lack of rigorous and long-term evaluations that isolate the re-employment outcomes for displaced workers. Beyond evaluations of individual programmes within specific countries, there remains a gap of targeted and longitudinal evidence to inform generalised conclusions about what works to support displaced workers.

Card et al. [96] produced arguably the most comprehensive evaluation of active labour market programmes (ALMPs). The meta-analysis conducted econometric evaluations for 207 recent ALMPs in Western Europe and North America. The study measured the impacts of ALMPs on the probability of obtaining employment. It did so by measuring the effects of specific ALMPs on particular populations at a given post-programme time horizon. The research reaches four substantive conclusions: [96]

1. The average short-term effects (less than a year post-program) of ALMPs did not yield statistically significant outcomes; however, the medium-term (1-2 years post-program) and long-term (2+ years post-program) effects both yielded positive and statistically significant outcomes, increasing over the time horizon;

2. The time profile of impacts varies according to the type of ALMP. For example, 'Job Search Programmes' yield relatively high short-term effects, 'Training and Private Sector Employment Programmes' have lower short-term effects but higher long-term effects, and 'Public Sector Subsidies' have negligible effects across all time horizons;

3. The average impacts of ALMPs vary across groups and that different ALMPs perform better for particular groups. For example, the average effects of ALMPs are larger for females and long-term unemployed than for older workers and youth. Additionally, Job Search assistance appears to be more successful for disadvantaged job seekers, whereas private sector employment subsidies tend to have higher outcomes for the long-term unemployed; and

4. ALMPs generally have larger impacts in periods during slower growth and higher unemployment.

Again, while the results from this comprehensive study provide general insights on ALMPs, it is limited with regards to displaced workers. Therefore, it is challenging to glean concrete conclusions on the efficacy of specific labour market programmes for displaced workers. This highlights a gap in the literature and should be prioritised for future research.

Labour displacement in Australia. Accurately measuring the incidence of labour displacement in Australia is difficult. This is largely because Centrelink does not differentiate retrenched workers as a distinct group of the unemployed and the ABS do not longitudinally measure employment outcomes for displaced workers [255]. This is also further complicated by the ABS definitional changes of retrenched workers in 2014, limiting the range of consistent data that is available.

Nonetheless, there are some measures of labour displacement in Australia that are collected by periodic surveys. The ABS provides quarterly estimates of retrenched workers [35] and the Household, Income and Labour Dynamics in Australia (HILDA) also periodically publish retrenchment rates [130].

Job turnover in Australia is relatively high, compared to other advanced economies in the OECD [308]. However, Australia's labour market is relatively robust and dynamic, so the majority of workers have been able to readily find new employment opportunities. According to analysis by the OECD of the HILDA data in 2011-12, 19.4% of Australian workers separated from their jobs. This is higher than the average separation rate of 16.6% among OECD nations [255]. According to HILDA, one-fifth of workers separated from their jobs were displaced due to economic reasons during the 2002-13 period [255]. This retrenchment rate of separated workers increased to almost 35% in 2009 as a result of the Global Financial Crisis, but then varied between 20% and 28% until 2013 [130, 255].

The ABS measure retrenched workers as a percentage of the total workforce [35]. The ABS disaggregates retrenched workers according to those currently unemployed, currently employment (that is, re-entered employment during the quarter), and not in the labour force (NILF). As seen in Fig. 2.6, the blue stacked bar charts show the stock of retrenched workers in the previous quarter and the shades of blue indicate their employment status for the current quarter: 'employed', 'unemployed', or 'not in labour force' (NILF). The orange line highlights the retrenchment rate as a percentage of the total Australian workforce. As observed, the total stock and rate of retrenched works has declined since 2014 to the last available quarter of data (February 2020, as of writing) [35]. However, given the size and severity of the COVID-19 shock on the Australian labour market, it is highly likely that the stock and rate of Australian retrenched workers has increased in the later quarters of 2020.

How labour market programmes support displaced workers in Australia. The majority of displaced workers in Australia do not receive targeted support. It is estimated that 50%, or more, of displaced workers are ineligible for income assistance, at least for an extended period following their retrenchment [255]. This is due to strict income and assets means testing. Instead, almost all displaced workers are assessed

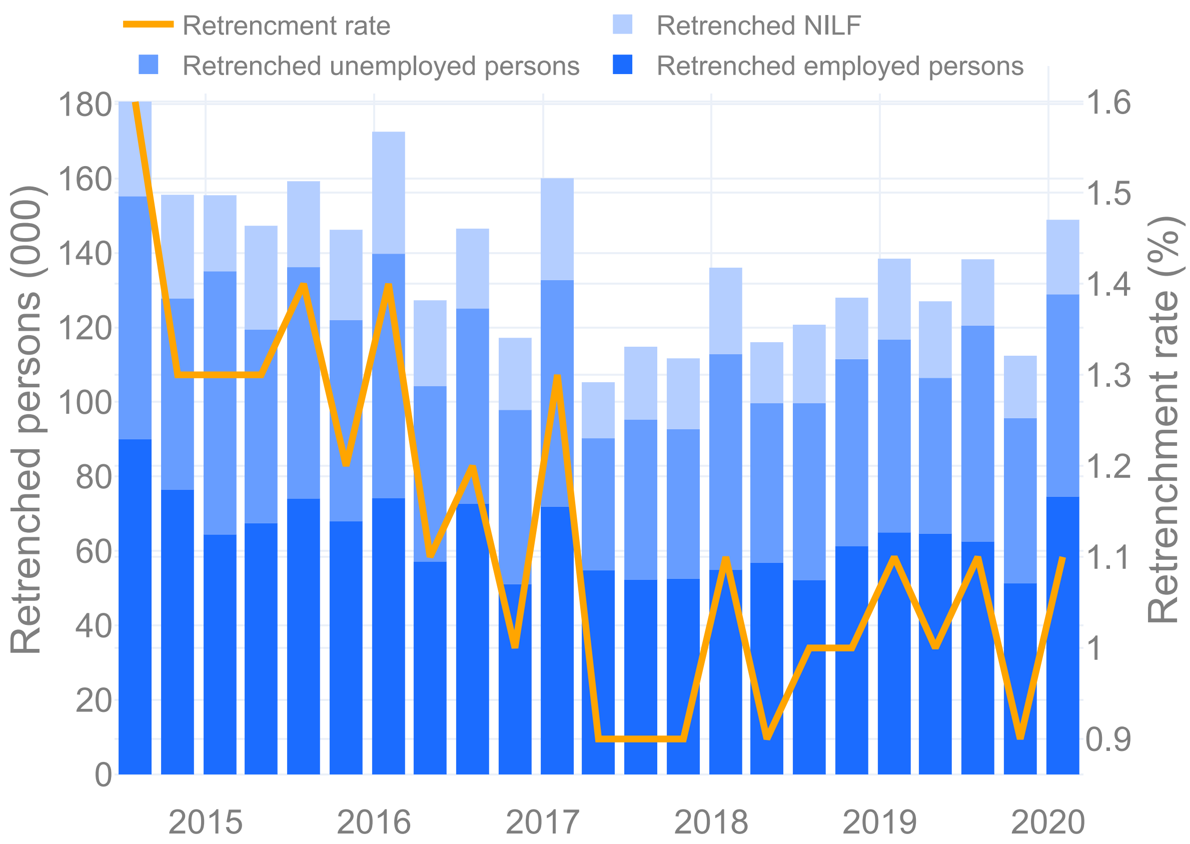


Figure 2.6: The number of persons (000's) who ceased a job during the last three months because they were retrenched or were self-employed whose business closed down. The stacked bar charts further indicate the current employment status of persons retrenched as 'employed', 'unemployed', and 'not in labour force' (NILF). The orange line highlights the retrenchment rate as a percentage of the total workforce.

and supported as part of the general unemployed population. While there are Structural Adjustment Programmes, such as the Automotive Structural Adjustment Programme, these industry-specific programmes provide limited coverage of the displaced worker population [255]. According to the OECD, less than 1% of employees in Australia's total 'dependent employment' population is covered by Structural Adjustment Programmes [255]. Therefore, the main forms of public assistance for displaced workers are those that are generally available to the unemployed.

There are two main support measures available to eligible displaced workers: (1) income assistance (passive measures); and (2) re-employment services (active measures). These labour market programmes are mainly funded by the Commonwealth Government of Australia, with exception of select co-funding arrangements with States for specific programmes [61]. A brief description of the two main support measures follows.

Income assistance in Australia. Unlike most OECD countries, Australia does not have an unemployment insurance scheme [254]. Instead, there are two main unemployment income support programmes:

1. *Newstart Allowance*: for unemployed people aged 22 and over; and

2. *Youth Allowance*: for unemployed people aged 16 to 21.

Both schemes are subject to strict means testing, and it is compulsory for recipients to undertake a government approved activity to lift their employment prospects (for example, training or counselling). The four means tests that must be satisfied in order to access NSA benefits are: (1) income; (2) assets; (3) liquid assets; and (4) the income maintenance period (for example, redundancy payments can disqualify or delay NSA payments) [40]. However, the means testing of employment services support policies have temporarily changed since the outbreak of COVID-19 [129].

Prior to the COVID-19 employment support policies, the means testing in Australia were relatively strict in comparison to other advanced economies [254]. This places significant limits on displaced workers accessing income benefits. Such limits can be problematic as being displaced from employment increases the probability of experiencing poverty in Australia [255].

Re-employment services in Australia. Australia is the only country in the OECD to provide fully privatised employment services [254]. 'Jobactive' is a network of private employment contractors across 1,700 locations and represents Australia's main ALMP [128]. These private employment service providers (ESPs) include a mixture of non-profit and for-profit ESPs that are contracted by the Commonwealth Government over three-year periods. ESPs have face-to-face consultations with job seekers to help structure a tailored 'Job Plan' of activities (such as CV writing, interviewing skills, job search etc.) to obtain employment. Additionally, ESPs draw connections with local employers to help understand their recruitment needs and to match employers with suitable job seekers. There are wage subsidies available to encourage employers to hire at-risk job seekers (Indigenous, long-term unemployed, and mature age job seekers) and ESPs have access to a discretionary 'Employment Fund' to assist with job seekers with personalised support (for example, specific training and work-related items) [128].

ESPs are evaluated quarterly on a five-star rating according to the employment and education outcomes achieved by their job seeking clients [128]. ESPs are remunerated through service fees and outcome-related performance indicators [255]. Prior to receiving jobactive services, Centrelink classified job seekers with the 'Job Seeker Classification

Instrument' into one of three 'Streams' to identify job-readiness and disadvantage; 'Stream A' job seekers are the most job ready, 'Stream B' job seekers greater levels of support, and 'Stream C' job seekers need the most support [128, 255]. The less job ready a job seeker is classified, the greater the service and performance fees available to ESPs. Displaced workers are most likely to be assigned as 'job ready' in Stream A and thus receive limited support in the first year of unemployment [255].

The two main goals of jobactive are: [128]

1. To make the re-employment system more 'activating' (that is, direct more job seekers toward finding employment and away from the income support rolls); and
2. To reduce the deadweight costs associated with paying fees to ESPs for assisting job seekers that likely would have found work without assistance.

Re-employment outcomes for displaced workers. As Centrelink does not differentiate displaced workers from the unemployed, it is difficult to isolate the re-employment outcomes of displaced workers. Nonetheless, the OECD provides estimates based on OECD data and data from the Household, Income and Labour Dynamics Australia (HILDA) Survey in 2013 [255].

According to OECD analysis in 2012-13, Australia performs relatively well compared to other OECD economies at transitioning displaced workers into re-employment within one-year [254, 255]. Approximately 70% displaced workers in Australia transition to re-employment within one-year, and approximately 80% within two years [255]. Re-employment rates, on average, continue to increase slightly in the third year, and then drop in the fourth year. Re-employment rates for displaced workers tend to be slightly lower than other workers experiencing other types of separation [255]. While the data is not relatively recent, and some country recordings are incomplete, it still provides insight into general trends.

Issues with Australian labour market programmes for displaced workers. The OECD have identified three main systemic issues with how displaced workers are measured and supported [255]. These include:

- *Limited support*: Re-employment support to displaced workers is limited in the first year of unemployment due to income and assets means testing;
- *Measurement problems*: Centrelink does not track displaced workers and their employment outcomes, so displacement levels are estimated based on periodic surveys; and

- *Low levels of training*: Displaced workers eligible for re-employment assistance receive relatively little training support in comparison to other OECD nations.

Clearly, redeploying displaced workers back into the workforce has benefits for individuals, firms, and the aggregate economy. However, moving between jobs can be challenging and, in some cases infeasible. There are multiple barriers that can impede successful transitions to new jobs, such as skill gaps, education requirements, experience demands, and other factors [57, 243, 244]. The following section reviews the literature related to job transitions and 'skill gaps' that can be a major impediment to labour mobility.

2.2 Job transitions & skill mismatches

This section discusses the factors affecting labour mobility, the causes and effects of skill mismatches, measuring human capital transferability, and accounting for the asymmetries between jobs.

2.2.1 Occupational mobility and job transitions

The related literature on job transitions broadly falls into the categories of labour mobility and human capital within the discipline of labour economics. 'Labour mobility' refers to the allocation of workers to firms and their ability to move between jobs [74]. Labour mobility is an important determinant of healthy labour markets. Efficient labour movements enable firms to hire more productive workers, effectively match workers to jobs based on their preferences, and helps to protect markets against economic shocks and structural changes. 'Human capital' refers to the skills, knowledge, capabilities, and experiences possessed by an individual that influence their productive capacities and that can be exchanged for labour at a prevailing market wage [58, 59, 296].

The process of labour mobility is constantly evolving and influenced by a variety of factors. These include labour market policies that shape hiring practices, job separation support programs, and relocation incentives [52, 178, 281]. It is also impacted by the extent of human capital in a labour market, which refers to the supply of skills, knowledge, and abilities of a labour force that firms can employ to produce goods and services [163]. According to Nedelkoska and Frank, skills should be considered part of human capital that is acquired through education, training, and work experiences [243]. While access to education and training undoubtedly affects the acquisition of skills, particularly 'general' skills [58], skills are also acquired through work experience. Typically, firm or industry-specific skills are not perfectly mobile across employers and can hinder labour mobility [335]. Therefore, the extent of human capital 'specificity' in labour markets is an important factor affecting labour mobility. It impacts how transferable skill sets are between jobs and reveals their underlying mismatch. The remainder of this section reviews the literature relating to the cause, effects, and measurement of skill mismatch and skill transferability. There are subtle, but important, differences between these terms. 'Skill mismatch' refers to the differences between the supply of and demand for skills in a labour market. Whereas 'skill transferability' is the capacity to leverage previously acquired skills to perform tasks across different jobs, either because the tasks are similar or the skills can be flexibly applied to different tasks [243]. The following literature forms the theoretical basis that directly informs our novel approach to measuring the distance between skills and sets of skills.

2.2.1.1 Causes and effects of skill mismatches

Skills provide the means for workers to complete tasks that are required by jobs. A distinction should also be made between skills, knowledge areas, and abilities. 'Skills' are the proficiencies developed through training and/or experience [263]; 'knowledge' is the theoretical and/or practical understanding of an area; and 'ability' is the competency to achieve a task [157], where a task is a unit of work required by a job. For simplicity, the term 'skill' will collectively represent these three definitions throughout this thesis.

Skill mismatch occurs when the skill demands of a job differ from the supply of skills available in a labour market [243]. When labour demand outweighs the supply of specific skills, it is referred to as a 'skill shortage'; when supply exceeds labour demand, it is often called 'skill excess' or 'over-supply'. Skill mismatches are closely monitored by labour economists due to the cost burdens they impose. For workers, they lower wages and employment opportunities; for firms, they restrict access to talent to implement specific tasks; for economies, they drain productivity [243].

The causes of skill mismatches can be both frictional and structural. There are search costs associated with a worker finding a job suitable to their skills, education, and experiences [321]. The frictions of matching workers to appropriate jobs can hamper efficient job transitions and exacerbate skill mismatches. Structural factors causing skill mismatches, however, are concerned with the supply and demand for skills.

Regarding the supply side, the literature has mainly focused on the role of public institutions to facilitate skill development through education and training. Freeman, among others, demonstrated that the oversupply of skills can depress wage premiums that are typically earned by highly educated workers [149]. Further, Goldin and Katz showed that the wage premium of US college graduates was a function of the supply of college education, with college wage premiums increasing when the supply of college degrees in the labour market was low [164].

Structural changes in the demand for skills are predominantly caused by (1) technological advances and (2) globalisation or trade. Concerning technology and innovation, Vona and Consoli [331] present a useful framework for understanding the evolving relationship between skills and technological change. In the early stages of new technology adoption, the authors argue that tasks are typically complex and non-routine. Consequently, specialised and highly skilled labour is required to make productive use of these new technologies. As time progresses, however, knowledge becomes structured and codified, enabling tasks to be routinised and automated. Eventually, the marginal benefits of specialisation diminish as the use of the technology becomes standardised and tasks are able to be performed by lower-skilled workers. Related, is the theory of **Skill-Biased Technological Change** (SBTC). The SBTC hypothesis posits that technologies disproportionately advantage highly-skilled labour over lower-skilled labour, as technologies tend to enhance the skills highly-skilled workers and automate lower-skilled workers [45, 66, 235]. The SBTC hypothesis was later modified to account for the relationship between computers and task-specific requirements of jobs. This Task-biased Technological Change (TBTC) framework [7, 44] classifies labour tasks along two main spectra; routine to non-routine tasks and cognitive to manual tasks. Computerisation, according to TBTC, tends to assist non-routine tasks and automate routine tasks. As a result, computers automate the labour tasks of middle-skilled workers (typically, routine-cognitive workers), which helps to account for recent dynamics such as declining real wages of middle-skilled workers and labour polarisation [167]. However, the argument that the negative demand-side effects of computerisation are limited to routine tasks is now coming under scrutiny. The rapid advances and diffusion of AI technologies cast doubt over this assumption. Brynjolfsson and McAfee [87] and Frey and Osborne [151] present compelling arguments that the automation capabilities of AI are extending to non-routine tasks, both in the cognitive and manual domains. Non-routine tasks that were previously considered out of reach by AI are quickly outperforming human levels in a range of non-routine tasks, such as Natural Language Processing (NLP) [82], Image Recognition [322], and unstructured learning tasks [309].

Globalisation or trade also has important implications on the demand for skills, which can exacerbate skill mismatches. Offshoring enables firms to fulfill their required labour tasks without personal contact, which can be managed electronically without a loss in quality [71]. This shifts the demand for skills towards countries with lower labour costs. Autor et al. [43] found that approximately one quarter of the decline in US manufacturing employment can be attributed to increasing trade with China.

Taken collectively, changes in the supply of and demand for skills alter the extent of skill mismatches in a labour market. This equilibrium is dynamic and directly affects labour mobility. The following subsection reviews the literature for measuring skill mismatch and transferability, which directly informs the research in Chapter 6.

2.2.1.2 Measuring skill transferability and mismatch

This research is part of a small but growing area of labour economics that measures the 'distance' between skills, jobs, and other defined skill sets. Among the earliest work in this area was conducted by Shaw [302, 303] who defined measures of occupational distances via proxies of skill transferability across occupations. This was based under the assumption that occupations with high levels of skill transferability are strongly correlated with high probabilities of transitioning between these occupations. This is an assumption that we adapt, test, and prove in Chapter 6.

More recent studies have made use of skill and task-level data, such as the US Dictionary of Occupational Titles (DOT - a predecessor to O*NET) or the German Qualification and Career Survey (QCS). Poletaev and Robinson [280] use task-level data from DOT to study the similarity between occupations. The authors construct four measures of basic skills, applying the factor analysis method used by Ingram and Neumann [189]. These four skill measures characterise the 'skill portfolios' of occupations, which are organised as vectors of skills. They then use Euclidean distance to compute the similarity between occupational skill vectors in order to identify which workers change their skill portfolios when transitioning between jobs. The authors show that workers who find jobs with similar skill requirements to their earlier jobs before displacement avoid large wage losses.

Similarly, Gathmann and Schonberg [158] use the QCS to classify occupations into a 19-dimension skill space defined by the survey. Each occupation represents a skill vector, where occupations consist of certain skills with varying degrees of mastery. The authors use the angular distance between the 19 skill vectors to position the occupations and measure their relative distances. The authors demonstrate that individuals transition to occupations with similar task requirements and that the distance requirements decline with greater work experience.

Most recently, Alabdulkareem et al. [14] used techniques from Network Science and unsupervised Machine Learning to illustrate occupational polarisation based on their underlying skill. Data sources included a combination of O*NET skill-level data and US occupational transitions data in the Current Population Survey from the US Bureau of Labor Statistics. The authors implemented an established measure from Trade Economics, called 'Revealed Comparative Advantage' (RCA), to firstly measure the relative importance of a skill in a job while normalising for high-occurring skills. After setting a threshold for skill importance, skill similarity was then calculated as the minimum of conditional probabilities that a skill pair are both important in a job when they co-occur. The authors then used these pairwise skill similarities to map workplace skills as a network, highlighting skill polarisation and proving a correlation with wage polarisation. Dawson et al. [120] extended this approach by applying this method to real-time job ads data to adaptively select occupations based on their underlying skill demands. This enabled the authors to accurately monitor changing labour demands and detect skill shortages for an evolving set of Data Science and Analytics occupations in Australia. The skill similarity methods applied by Alabdulkareem et al. [14] and Dawson et al. [120] provide the foundation for the skills set similarity methodology in Chapter 6.

While all of these approaches represent significant contributions in the evolution of measuring skill transferability, there is one major shortcoming. All of these methods yield symmetric distance measures. That is, the distance from one skill or occupation to another is the same despite the direction. For example, according to these methods, it is just as difficult for a Nurse to become a Surgeon as the other way around. Intuitively, however, acquiring certain skills to transition to a particular occupation is more difficult in one direction than the other. In this sense, skill acquisition and occupational transitions are directed and asymmetrical.

Asymmetric measures for skill mismatches. Nedelkoska et al. [244] develop skill mismatch metrics that account for the strong asymmetries in the transferability between skills. The authors construct occupational skill profiles by using factor analysis to extract five task-based skills on German administrative and data on individuals' work histories. They then calculate the share of workers in each occupation carrying out these tasks. The average years of education and training associated with each task are used as weights to indicate skill complexity required by an occupation. Adding these weights reveals asymmetries between skills and therefore occupations. They show that by switching occupations, people incur both skill shortages and skill redundancies, which results in significant wages losses up to 15 years following the job displacement. While accounting for skill asymmetries represents a clear improvement for measuring the distance between skills, using years of education and training as the sole proxy for skill complexity is questionable. As previously stated, work experiences are an important contributor to the acquisition of skills and causes of mismatch.

Bechichi et al. [57] adapt the Nedelkoska et al. [244] model by analysing occupational data from the OECD Survey of Adult Skills (PIAAC). They firstly use the six task-based skill indicators from PIACC [169]. The authors then measure these indicators on 127 occupations (at the 3-digit occupational level) across 31 different countries to assess occupational distances based on 'cognitive skills' and skills acquired from tasks 'on the job'. This method accounts for skill asymmetries and skills acquired from work experiences. The resulting 'skill shortage' and 'skill excess' measures are then used to predict education and training resources required to transition workers from one occupation to another. Therefore, this research represents another advance toward the goal of accurately measuring skill and occupational distances. However, a minor shortcoming of this work is that it is performed at the 3-digit occupational level, which is a relatively high classification level (1-digit being the highest and 6-digit being the lowest and most detailed). Additionally, surveys provide lagging data that are typically slow to report and expensive to conduct. This is problematic in labour crises, such as the job displacements caused by COVID-19. Dynamics of labour markets quickly change in times of crisis and displaced workers are faced with transitioning between jobs with rapidly evolving skill demands. Real-time data, therefore, becomes essential.

The research presented in Chapter 6 builds on these significant works and addresses both of these shortcomings by using real-time job ads data and applying a method capable of measuring the distance between any defined set of skills, such as occupations at the detailed 6-digit occupational level, industries, or even personalised skill sets.

2.3 AI & its impacts on labour markets

Technological progress is a main driver of economic growth and improved standards of living [159]. Throughout history, technologies have increased productivity and aggregate demand, thus lifting per capita income, wealth, and quality of life [237]. This progress has mostly been incremental with established technologies improving over time and diffusing throughout society. In key moments of history, however, technological progress has been 'transformative', creating pervasive structural changes to economies and societies [150].

These breakthroughs resulted from a confluence of technological advances, economic conditions, and socio-political dynamics that have enabled widespread adoption and diffusion [84].

Over the past three centuries, there have been three major industrial revolutions resulting from technological progress [299]. The first industrial revolution, broadly spanning from 1760 to 1840, was characterised by the invention of the steam engine and the growing use of machines automating manual labour in manufacturing. The second industrial revolution, starting in the late 19th century until the early 20th century, was a consequence of electrification and the assembly line, which made mass production possible. And the third industrial revolution, beginning in the 1960's until the 1990's, is often referred to as the 'Computer' or 'Digital' revolution ('computerisation' henceforth). This era saw the advent of semiconductors and mainframes that enabled the proliferation of personal computers and the exponential growth of information processing. It also marked the rise of the Internet, which digitised access to information and communications for billions of people.

The recent improvements in AI have been a source of great optimism, leading to scholars and experts to pronounce the beginnings of the 'Fourth Industrial Revolution' [17, 77, 87, 206, 299]. To an extent, this reflects the continuing progress spawned by computerisation from the last industrial revolution. In contrast, the autonomous capabilities of AI technologies, particularly Machine Learning, represents a fundamental departure from the first wave of computerisation [89].

This section will provide a definition of AI and its subset technologies, a discussion of how AI differs from computerisation, and an overview of the arguments positioning AI as the next major General Purpose Technology and its projected socio-economic impacts.

2.3.1 Defining AI

The term 'Artificial Intelligence' has been credited to John McCarthy, who organised a Summer Research Workshop on Artificial Intelligence at Dartmouth College in 1956 [229]. The purpose of convening the workshop was to investigate the ways that machines could emulate aspects of intelligence. The workshop unified the field and provided a platform to incorporate important earlier technical ideas that have come to characterise AI today. These include statistics, probability [54], logical reasoning [73], electronic computing [306], and previous developments in computer science, such as the 'Turing Test' developed by Alan Turing [324].

AI is notoriously difficult to define due to the conceptual ambiguities of intelligence. As McCarthy states, any definitions of intelligence usually relate to human intelligence because we cannot generalise the computational procedures we consider 'intelligent' [228].

Russell and Norvig organise definitions into four categories: (1) thinking humanly; (2) acting humanly; (3) thinking rationally; and (4) acting rationally [292]. The authors place a strong emphasis on the 'rational agent' in their definition, whereby such an agent (that is, an AI system) is goal-oriented and "acts [autonomously] so as to achieve the best outcome or, when there is uncertainty, the best expected outcome" [292].

From a policy and regulatory perspective, this 'rational' and 'goal-oriented' definition is not always informative. Such a definition simply replaces the abstract term of 'intelligence' with the similarly difficult to define term of 'goal' [295].

To move beyond semantics, this research adopts a broad view on intelligence as "the quality that enables an entity to function appropriately and with foresight in its environment" [246]. Therefore, AI can be considered an umbrella term, broadly defined as a group of technologies that are capable of performing tasks autonomously, which if performed by a human, would be considered to require intelligence [17, 112, 295].

2.3.2 Categorising the Extent of AI

Classifications of AI also differ according to the capabilities and strengths that AI systems possess. The extent of AI can be classified in the following ways:

- *Narrow Artificial Intelligence*: AI that performs well on specialised tasks in specific domains but does not transfer to other areas [327]. Narrow AI is a present-day reality with examples such as financial fraud detection systems and speech recognition software.

- *Artificial General Intelligence (AGI)*: Sometimes referred to as 'Strong AI' or 'Human-level AI', AGI refers to an AI agent that is at least as smart as humans across many or all domains [269]. Significant AI research and development is directed towards AGI and is the explicit goal of several leading AI research and development organisations [286].

- *Superintelligence*: Bostrom defines Superintelligence as an intellect that greatly exceeds human intelligence in practically all forms [77].

The greater extent of AI corresponds with a longer time-horizon to its possible attainment. However, predictions of time-horizons are fraught with uncertainties. A survey of leading researchers in AI found that predictions for when AGI would likely be achieved varied from 10-70+ years [240]. This, however, remains highly speculative. Therefore, this research focuses exclusively on the Narrow AI technologies that are present now and likely to grow over the coming decades.

2.3.3 Research Trends & Subsets of AI

There are a host of subsets and capabilities of AI that aim to automate or replicate intelligent behaviours. Russell and Norvig identify the following as among the most significant capabilities [292]:

- *Logical and automated reasoning*: To use stored information to answer questions and draw new conclusions based on formal logic.
- *Knowledge representation*: Storing and representing information about the real world in forms that machines can use to solve problems.
- *Planning and navigation*: Representing sequences of actions and world models, reasoning about the effects of actions, and techniques to efficiently navigate the space of possible plans.
- *Natural language processing*: Ability to analyse, synthesise, and represent human language with machines.
- *Perception*: Sensing and perceiving the environment by acquiring, interpreting, selecting, and organising information.
- *Robotics*: Interact with the environment, manipulate objects, and move in the physical world.
- *General intelligence*: Combining the discrete parts of intelligence to create an autonomous intelligent agent that can successfully generalise across tasks and domains.

Capabilities and techniques are often combined to achieve tasks requiring intelligence. An elaboration of Machine Learning (the most dominant sub-field of AI) follows.

2.3.3.1 Machine Learning

Machine Learning (ML) is currently the preeminent subset of AI, in which the greatest progress has been made. Its omnipresence and growing use cases have largely stemmed from its abilities to lower the price of prediction [13].

ML seeks to allow machines to learn on their own and improve with experience. Unlike Expert Systems that are explicitly programmed to anticipate the desired response to all possible inputs, ML systems 'learn' from data to adapt and autonomously make predictions and decisions [212]. The learning process is based on probability and statistics, where ML algorithms are 'trained' on input data to infer probabilistic predictions. Once trained, ML algorithms are then 'tested' with data inputs outside the examples observed in the dataset [292]. If the testing performance of the algorithm is robust, and assuming the training data and environment are representative, then the designers could decide that the ML system is ready for deployment.

The three main algorithmic techniques of ML are Supervised, Unsupervised, and Reinforcement Learning [192]. Additionally, Deep Learning and Artificial Neural Networks provide algorithmic layers and infrastructures to autonomously find representations in masses of data and self-improve.

Supervised Learning, the most widely used ML technique, uses data overseen by a human expert (the 'supervisor') to train an algorithm to predict and/or make decisions based on data that is correctly organised [315]. The purpose of Supervised Learning is to learn by analysing these vast reams of labelled data (that is, organised data) to find representations, or generalised trends. Once designers are confident that the ML system is 'trained' to robustly categorise situations and make predictions, the ML system is then 'tested' on situations not present in the training set. If the ML system performs well during testing, designers may apply the system in real-world environments.

The two main tasks of Supervised Learning are:[292]

- *Classification*: Correctly assign class labels to unseen instances. For example, correctly identify cats in YouTube videos.
- *Regression*: Predict continuous values based on inputs. For example, predicting house prices based on metrics, such as location, square footage, number of bedrooms, and other information.

Unsupervised learning attempts to autonomously find structure and patterns hidden in unlabelled data. Two major methods of unsupervised learning include: [292]

- *Clustering*: Organising data into groups based on similarities and relationships. For example, identifying similar occupational groups according to their underlying skill demands.

- *Anomaly Detection & Association*: Identifying the outliers to reduce the complexity of data, while keeping a relevant structure as much as possible. For example, detecting credit card fraud from large and unstructured financial transactions data.

Reinforcement learning (RL) is a technique where an agent learns through trial and error, and by being rewarded for success [315]. The RL agent does this through 'trial and error', receiving feedback on the amount of reward that a particular action yields. By analogy to a child's game, this is like a blindfolded child being told 'hot' or 'cold' by an observer as they get closer or farther away from an unseen target object.

Unlike in Supervised Learning, an RL agent is not trained on labelled examples around the correct actions to take. RL is also different from Unsupervised Learning because it is not trying to find a hidden structure in unlabelled data. While uncovering patterns and relationships in data might be helpful to a learning agent, and there are examples of combining these two approaches, RL is principally concerned with maximising its reward function. An RL agent learns from direct interaction with an environment, without relying on complete models of an environment or strong supervision. For example, Google Deepmind's AlphaGo Zero used RL techniques to become the strongest Go player in history [177].

Of all the forms of ML, RL represents the closest form to how humans learn, and RL is predicted to play a crucial role in the quest for General Artificial Intelligence [17].

Deep Learning & Artificial Neural Networks. Deep Learning (DL) is an AI method that uses specialised computing infrastructures called Artificial Neural Networks (ANNs) to process information in ways inspired by the human brain [165]. ANNs learn from masses of data examples and they cannot be programmed to perform specific tasks by following an algorithmic approach (as seen in the previous types of ML). Instead, ANNs are fed data to learn by example, identify patterns in data, and autonomously make sense of complex information by finding representations (that is, common features), with the goal of becoming more accurate over time.

A major advantage of DL is that it can use ANNs to find patterns and solve problems that humans do not always know how to do. For example, natural language has a huge variance in vocabulary within the same language. DL can be an effective means to learn function approximators to make sense of this complex information, such as correctly interpreting accents. The disadvantage, however, is that because DL autonomously uses ANNs to solve problems, its operation and results can be unpredictable and difficult to interpret.

ANNs are made up of different layers of nodes, as represented by the vertical rows of circles in Fig. 2.7 [134] (analogous to neurons). These nodes are connected to the next layer of nodes, seen in Fig. 2.7 by the connecting arrows. The nodes and connections have a variable 'weight', which is a number between 0 and 1, to signal the strength of particular connections. These weights are said to 'transform' the input data into composite pieces, so that it can start to recognise patterns [165]. If during training the ANN does not correctly identify a pattern, then the weightings adjust. By the time the data is fed through the network and reaches the final layers, the ANN will have created a complex system of detecting features (that is, pattern representations) in the data. The output layer (green node in Fig. 2.7 then produces a 'label' or a value, which is the final output. After a period of 'training' with lots of examples, the ANN is optimised to autonomously identify features in the input data and correctly label new data that is fed into the system.

A simplified example of how ANNs operate in practice is when they are applied to recognise images. Consider an ANN being trained to recognise Cats by showing it lots of pictures of Cats. The images will need to be represented appropriately and there will need to be a way of providing feedback that compares the output that the network produces with the output it should have produced. The difference will then transform the weights of connections between the units in the network. Rather than examining the images of Cats in their entirety and instantly responding, the neural network is trained to identify features of different Cats. Patterns are identified between different data inputs usually through binary numbers, which is like answering yes/no questions (for example, Are there two eyes? Are there two ears? Is there fur?). Each neuron assigns a weighting to its received data, which is basically saying how 'correct' it is (or how strongly those nodes are connected). The final output is then determined by the total of these weightings, which is the probability of each picture correctly labelling a Cat.

2.3.4 Drivers of Recent AI Advancements

The recent breakneck advances of AI are being brought about through a confluence of developments. The driving factors are:

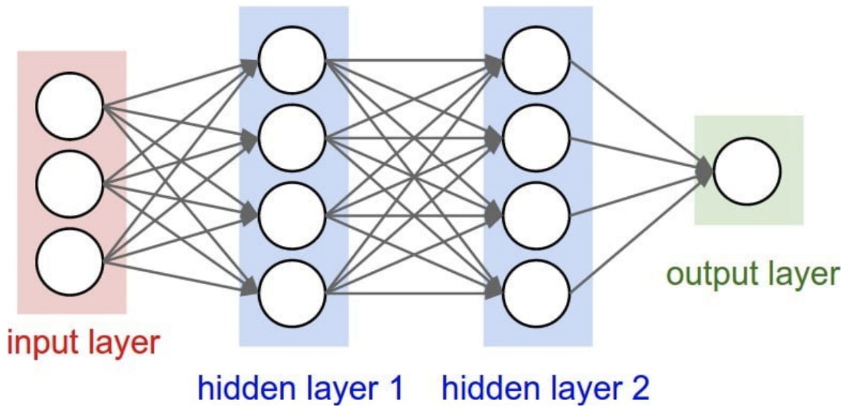


Figure 2.7: A basic illustration of a feed-forward neural network architecture used for DL.

- *Greater computational power and cloud-based infrastructure*: Until very recently, Moore's Law has held constant since 1971, with Central Processing Unit transistor counts doubling every two years [334]. The same applies for quality adjusted microprocessor prices, memory capacity, sensors, and pixels in digital cameras [114]. This increase in computational power also corresponds with plummeting barriers to entry with lower costs and easier access made possible by cloud-based infrastructure.
- *Increased quantities and access to training data*: The proliferation of devices and applications have exploded, which has caused the growth of data to follow a similar trajectory to Moore's Law. To put this in perspective, approximately 90% of data today has been created in the last two years [187]. This increase in data has led to more AI systems to be trained with more data. This growth, however, has disproportionately benefited select technology companies whose platforms collect this proprietary information.
- *Better algorithms and techniques*: As discussed above, research advances, particularly in ML, have led to more sophisticated algorithms and techniques in AI systems. This has enabled large-scale data processing and more effective applications of AI.
- *Growing investment*: Growth in private AI investment has increased with an average annual growth rate of over 48% between 2010 and 2018 [276].
- *Open source frameworks and libraries*: The open sourcing of AI tools such as Google's TensorFlow, Facebook's PyTorch, and Amazon's DSSTNE have made it

more accessible than ever for companies and individuals to apply their data to some of the most powerful AI technologies.

As these advances continue to develop, they will likely have an ever greater impact on more aspects of society. An exploration of AI as the next major General Purpose Technology and its projected economic impacts follows.

2.3.5 AI as the Next General Purpose Technology

The breadth and depth of capabilities that AI can perform have led several scholars to classify AI as the new General Purpose Technology (GPT) [8, 12, 89, 109, 151, 299, 323]. Bresnahan and Trajtenberg define GPTs as a technology, or class of technologies, with pervasive applications that affect the whole economy by transforming household life and the operations of firms [80]. The authors argue that GPTs are characterised by the following elements:

1. *Pervasiveness*: The GPT spreads to the majority of sectors.

2. *Improvement*: The GPT improves over time thus lowering the costs to its users.

3. *Spawn complementary innovations*: The GPT lowers the barriers to innovations and enables the invention of new products or processes.

Jovanovic and Rousseau add that the beginning of a 'GPT-era' is when the GPT has achieved a one-percent diffusion in the median sector [194]. Examples of GPTs include the steam engine, the internal combustion engine, electrification, and computerisation. While AI appears to be trending towards fulfilling these criteria [89], AI still remains a group of emerging technologies with regards to adoption and diffusion within the global economy [93].

2.3.5.1 Additional characteristics of General Purpose Techonologies

Historical examinations of specific GPTs, such as Electrification and Computerisation, have also identified less direct measures in identifying the evolution of GPTs, which include [194]:

1. *Slowing productivity*: The GPT may not be user-friendly in its early stages, so output falls initially and the economy adjusts.

2. *Rising skill premium*: If the GPT is not user-friendly, then the demand and earnings for skilled labour that can operate the technology rises disproportionately to that of unskilled labour.

3. *Market entrants, exits, mergers and acquisitions should rise*: Economic assets should be reallocated as the GPT is adopted and diffused.

4. *Stock prices should initially fall*: The value of capital that is being disrupted by the GPT should fall, depending on the speed of the GPT's arrival.

5. *Young firms should do better*: New firms will often bring the new products and processes of the GPT to market. The market share and value of new firms relative to old firms should therefore rise.

6. *Interest rates and trade deficits*: The rise in desired consumption relative to output should cause interest rates to rise or the trade balance to worsen.

These measures offer some support to the theory of AI emerging as the next GPT on a scale similar to Electrification and Computerisation. For instance, Brynjolfsson et al. argue that lags in diffusion are the primary reason why AI has not impacted measured productivity growth [89]. Indeed, productivity across the world has slowed, with 36 of 37 advanced economies had slower productivity growth in 2006-2016 compared to 1996-2006 [153], even before the COVID-19 induced economic crisis. Across these economies, productivity growth had slowed from 2.7% average growth rate to a 1% average growth rate. Brynjolfsson et al. argue that this lull in productivity is consistent with the GPT theory that productivity is subdued in the emerging stages of a new GPT, before dramatically increasing as the economy adjusts [89].

AI wage premiums are also rising disproportionately relative to average wages. In Australia, occupations that require high-levels of AI skills, such as 'Data Scientists' and 'Data Analysts', earn a wage premium approximately AU$30,000 above median wages [120]. This reflects the high demand for skilled labour in AI. It also highlights the low-levels of standardisation and user-friendliness to operate AI systems, thus requiring highly specialised skills, which is similar to the computerisation era of the 1990's.

While the GPT theory is not absolute, it provides a framework to compare AI with other transformational technologies, such as Electrification and Computerisation, to help understand AI is adopted and diffused in the economy.

2.3.5.2 Differentiating AI from Computerisation

AI represents a potential departure from other GPTs due to the scope of capabilities, the speed of development, and the scale of impact. Building upon the technological transformations of Computerisation and digital communications, AI is performing non-routine tasks that would otherwise require human cognition. Technological automation has traditionally occurred in areas of routine and manual labour because these tasks are relatively simple to codify [47]. AI expands the scope automation to include cognitive and non-routine tasks [151]. Machine Learning systems are being applied to trade on financial markets, diagnose diseases, and identify weather patterns. Tasks that have previously required human intelligence are being performed by AI applications at large scales and lightning speeds.

The multi-use capabilities of AI techniques have developed at an almost breakneck pace over the past two decades, and development continues to accelerate [276]. This positions the economic impact of AI to be one of the most significant in the history of GPTs [87]. Therefore, the implications of AI should be examined according to the degree of structural changes to national economies. Among the most important is the impact that AI will have on labour demand.

2.3.6 Projecting the Economic Impacts of AI

Scholars have put forward predictions and interpretations of how current AI technologies will impact the dynamics of future national economies. For instance, Makridakis draws parallels between AI and other technological transformations, highlighting the likelihood of profound productivity and output improvements, while also posing grave risks to economic inequalities between and within nations, firms, and individuals [221]. Similarly, Aghion et al. model the broadening scope of tasks that can be performed by AI, showing the potentially significant sectoral shifts and changes to firm organisational structures, which could lead to higher aggregate growth but constrained competition [11].

Among the most relevant to this research, however, is work completed by McKinsey Global Institute (MGI) on modeling the economic impacts of AI on the global economy [93]. This research analyses how AI technologies are adopted and diffused with reference to microeconomic behaviours of firms across various sectors. MGI conducted an econometrics simulation to project the economic disruptions that countries, companies, and individuals could experience as they transition to greater uses of AI. The simulation finds that by 2030 the number of firms that account for approximately 85% of economic activity will have adopted at least one AI technology, and firms that account for about 50% of economic activity will have adopted several. As the rates of adoption and diffusion accelerate, AI could deliver US$13 trillion of global economic activity by 2030, which is equivalent to an increase of 17% cumulative output from 2018.

MGI identifies the three most prominent factors affecting the way AI may affect economic growth as: [93]

1. *AI technologies substituting existing labour*, which could add 11% or US$9 trillion to global economic value by 2030.

2. *Product and services innovation and advances in global value chains*, which could add 10% or an output increase of US$8trillion by 2030.

3. *Disruption to firms and individuals*, which could cause negative externalities and transition costs that reduce total output by 9% or US$7 trillion by 2030.

As acknowledged by MGI, these dynamics depend significantly on assumed rates of adoption, diffusion, negative externalities (such as inequality and social stability) and labour market transitions – all of which are discussed further below in this Chapter and proceeding Chapters. Also, while high-level details are provided on the methodology, the granular details of the econometric model and input data have largely been withheld as proprietary information. This makes it difficult to assess the underlying validity of their modeled outputs. Nonetheless, the research offers useful insights into the potential economic impact of AI and the factors affecting its growth.

2.3.6.1 AI and Labour Automation

The advances of AI have led scholars to assess its effects on the scope, scale, and speed of labour automation. A useful starting point to examine the susceptibility of employment to technological automation (and AI automation, more specifically) is the 'Task Model', developed by Autor et al. [47]. In the Task Model, occupations are made up of a series of tasks that require particular skills. These skills and tasks are organised as being either routine or non-routine (vertical axis) and cognitive or manual (horizontal axis), which characterise the requirements for occupations. Fig. 2.8 provides some examples of occupations whose prevailing skills and tasks align to different quadrants of the Task Model matrix.

As discussed further in the 'Creative Destruction' section, routine tasks have the greatest susceptibility to technological automation [47]. Non-routine tasks (both cogni-

	Cognitive	Manual
Routine	Bookkeepers Bank teller Secretaries Travel agents	Professional divers Laundry cleaners Machine operators Product assemblers
Non-routine	Psychologists Photographers Architects Managers	Builders Electricians Plumbers Domestic cleaning

Figure 2.8: Examples of occupations according to the 'Task Model' put forward by Autor et al. [47].

tive and manual) that require critical-thinking, judgement, persuasion, creativity, and situational flexibility are much more difficult to automate [68]. Therefore, according to these research, occupations in the bottom two quadrants of Fig. 2.8 are less likely to be displaced by AI automation over the come two decades.

Research on the future rates of AI automation have built upon the Task Model, creating more complex frameworks. Among the most cited and influential has been the research of Frey and Osborne, who developed a methodology to assess the probability that computerisation will automate US occupations, with a focus on ML and Robotics [151]. Frey and Osborne ranked 702 detailed occupations on their susceptibility to automation according to specific task-level data from O*NET, an online service developed for the US Department of Labor. The authors organised their methodology using a combination of expert opinion and objective measures to assess occupational attributes, such as perception, manipulation, creativity, and social intelligence. From this analysis, the authors concluded that 47% of all jobs in US are potentially 'automatable' over the next 10-20 years [151].

Applying this methodology, other researchers have adapted this approach to assess automation susceptibility in other countries. The World Bank found that almost 60% of occupations in OECD countries are susceptible to technological automation over the

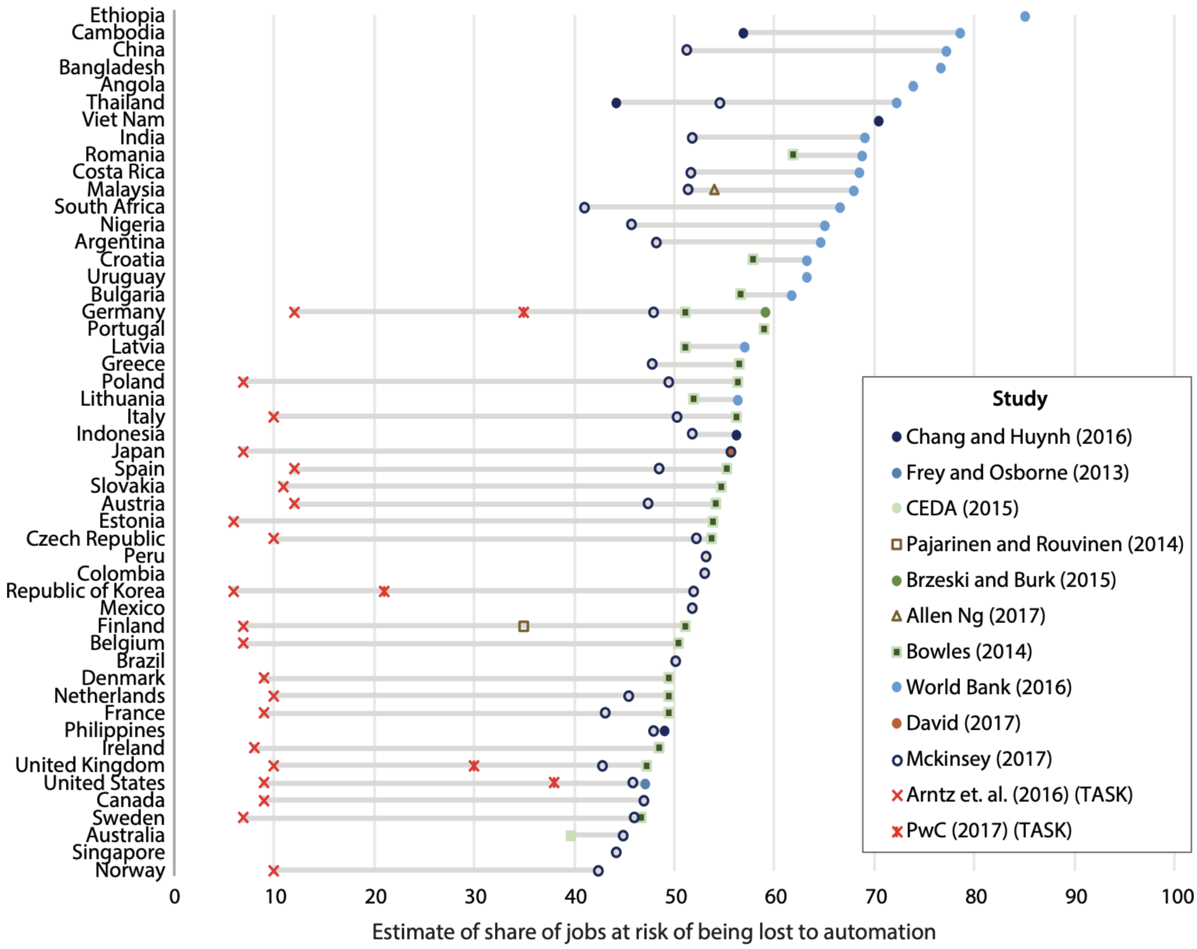


Figure 2.9: Comparison of labour automation risk by study as shown by Bruckner et al. [84].

coming two decades [236]. Fig. 2.9 highlights the range of job loss estimates throughout the world that are caused by automation from AI and other advanced technologies, by study and methodology [84].

In Australia, researchers that applied Frey and Osborne's methodology found that 44% of current occupations have a high probability of being automated by computerisation over the next two decades [136]. The researchers found that Labourers, Machine Operators and Drivers, Clerical and Administrative workers are among the most susceptible to automation because of the routine based nature of their occupations [136].

A limitation of the studies that employ Frey and Osborne's methodology is that they focus on the automation probabilities of whole occupations, rather than labour tasks and skills. Arntz el al. from the OECD challenge this approach [22], identifying that occupations consist of many specialist tasks that cannot be automated easily. The authors argue that assigning probabilities at an occupational-level likely leads to an overestimation of potential job losses by automation. Therefore, taking a task-based approach, Arntz el al. yield much lower estimates, predicting 9% on average across OECD countries over the next two decades [22].

2.3.7 Creative Destruction: Long-term relationship between technology and jobs

Technological progress is one the main long term drivers of aggregate economic growth and improvements in standards of living [68, 84]. Advances in technologies lift productivity, which increases per capita income, consumption, and aggregate demand.

As previously discussed, technological progress has mostly been incremental by making marginal improvements to existing technologies [237]. It takes time to refine and spread technologies throughout society. Occasionally, however, technological progress has been transformative and revolutionary, as seen during the Agricultural and Industrial Revolutions. Many scholars argue that the development of AI technologies will create a similar transformative era [87, 94, 148, 295].

Such technological revolutions can be the source of profound disruptions and anxieties. They can cause a process of 'creative destruction' that devalues incumbent markets and occupations, and replaces them with new opportunities for wealth [297]. In this sense, technologies change how jobs are performed and the number of people required to do them. This can cause significant short-run disruptions to labour markets [10]. While the long-run benefits have been significant, the prospect of rapid and systemic change fuels fear and anxiety [237]. These anxieties have reemerged as AI and other forms of computerisation continue their ascent in capability and prominence.

The most prominent fear is that AI will replace human labour on a large scale, resulting in mass unemployment or underemployment. As wage income from labour accounts for the largest source of income for the majority of people [279], AI automation of human labour threatens to reduce earning potential by displacing workers, which could exacerbate income inequality and social stability [206]. While this dynamic should be considered a serious risk, particularly in the context of rising income inequality in many major economies and labour's falling share of income relative to capital [279], these fears must be placed in the historical context of technological change.

2.3.7.1 Automation Anxiety

The awakening of the Industrial Revolution evoked fears of mass-unemployment by mechanical automation. In the 19th century, textile workers formed the 'Luddite Movement' to protest the automation of textile production. Workers damaged or destroyed machines and threatened the unravelling of social order.

Automation anxiety was also been prevalent in periods of economic stagnation. During the Great Depression, John Maynard Keynes voiced concerns on technological unemployment in his 1930 essay, 'Economic Possibilities for our Grandchildren':

> *"We are suffering, not from the rheumatics of old age, but from the growing-pains of over-rapid changes, from the painfulness of readjustment between one economic period and another. The increase of technical efficiency has been taking place faster than we can deal with the problem of labour absorption; the improvement in the standard of life has been a little too quick."* [203]

And similar concerns have been raised during periods of rapid economic growth and industriousness. Anxieties of technological unemployment in the 1950s and 1960s were severe enough that in 1964, President Lyndon Johnson created a 'Blue-Ribbon National Commission on Technology, Automation, and Economic Progress' [223]. The purpose of the commission was to assess whether rapid productivity growth would outstrip the demand for labour. The commission found that while technological change can cause structural shifts to the types of jobs demanded, there is no evidence that it eliminates the demand for labour:

> *"The basic fact is that technology eliminates jobs, not work."* [223]

These anxieties have been a constant throughout history. It echoes back to Greek Mythology as cautioned through the 'Promethean Myth' [135]: Man acquires fire from the Gods; fire becomes the source of Man's pain and suffering. We hear variations of the same narrative today: humans build and deploy AI; AI replaces our jobs and our purpose [173].

Yet, the fears of the 'Promethean Myth' have been consistently unfounded [42, 237]. To take the position that 'this time is different' is an abnegation of economic history.

Regardless, automation anxieties have re-emerged. As AI increases the scope of capabilities that can be automated by machines, concerns abound that many labour tasks performed by humans will be replaced. In their provoking book, *The Second*

Machine Age, scholars Erik Brynjolfsson and Andrew McAfee present a troubling view of the effects of labour automation:

> *"Technological progress is going to leave behind some people, perhaps even a lot of people, as it races ahead. As we'll demonstrate, there's never been a better time to be a worker with special skills or the right education, because these people can use technology to create and capture value. However, there's never been a worse time to be a worker with only 'ordinary' skills and abilities to offer, because computers, robots, and other digital technologies are acquiring these skills and abilities at an extraordinary rate."*

It is obvious, however, that the past two centuries have not rendered human labour obsolete. Global unemployment has trended downwards, and the employment-to-population ratio steadily rose during the 20th century [42]. In fact, a study performed by Deloitte found that new technologies over the past 144 years have created more jobs than they have destroyed [311].

This rebuttal to automation anxiety is not to dismiss the structural challenges that AI technologies pose to labour markets. Nor is it to place a blind faith in the determinism of history. The challenges of transitioning displaced workers to meet the new labour demands created by AI are likely to be significant. Potentially as profound as the structural adjustments experienced during the Agricultural and Industrial Revolutions [89].

However, history suggests that the demand for labour is not the concern. Instead, the concern is the speed of transitioning displaced workers to meet new labour demands. In other words, it is likely that old jobs will be replaced by new jobs, but it is the speed of transitioning workers to these new jobs that could be problematic.

A large part of this transition speed will depend on the types of jobs created and the skills demanded to fulfil these jobs.

2.3.7.2 Sector shifts, skills, and wages

As the world's most advanced economy, it is instructive to consider the sectoral employment shifts that have occurred in the United States (US) since the Industrial Revolution. Many of the same dynamics have been reflected in the Australian labour market. For simplicity, it is common to split labour markets into three main sectors: agriculture, industry, and services. Until around 1800, the majority of Americans were employed in the agricultural sector [156]. Just one century later, the employment rate in agriculture had halved. Significant portions of agricultural labour were automated, people

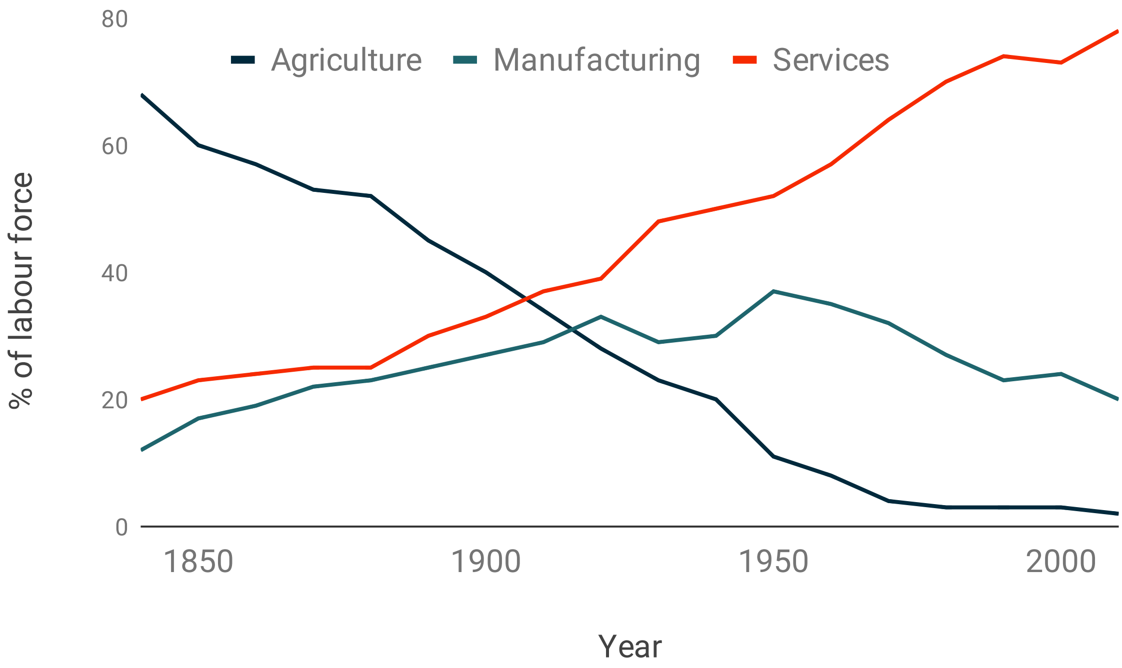


Figure 2.10: Distribution of labour force by agriculture, manufacturing, and services in the US from 1840-2010 [156].

moved to urban centres for industry-based jobs, and as education improved, a growing populace took up service professions. By 1900, the three sectors were almost evenly distributed [156]. This was a seismic shift from rural towns and farmlands, to urban cities and factories.

And the pace of change continued to accelerate. As Fig. 2.10. shows, by 2010, only 2% of Americans were employed in agriculture, 20% in manufacturing, and the 78% majority assumed service-based employment [55].

While the agricultural and manufacturing industries experienced job replacement by Information Technologies in the later decades, computer-intensive jobs in the service sector grew at much faster rates [68]. A major cause of these labour shifts is that agricultural and manufacturing sectors experienced greater increases in output per worker than services [53]. While computerisation has yielded productivity gains to many service sector industries [48], these service industries also tend to adopt more early-stage technologies and employment does not decrease with growing output per worker [68].

An explanation for this dynamic in the service sector is that computerisation is observed to complement the productivity of workers skilled in nonroutine tasks [46].

This is because nonroutine tasks are less predictable and more difficult to codify. And the productivity of workers adept in these skills is enhanced as routine tasks are more likely to be automated by computerisation. For example, a capable Data Analyst can quickly perform the routine tasks of sorting, processing, and visualising vast amounts of data, which leaves more time for the nonroutine tasks of interpreting and communicating insights from the analysis. Therefore, computerisation is said to be complementary to the productivity of workers skilled in nonroutine tasks [46].

These nonroutine skills, however, tend to be more complex and require greater periods of training and learning. This offers an explanation for the higher levels of labour demand and wage premiums commanded by higher skilled workers [68]. Many of the nonroutine skills of the service sector are classified as high-skilled tasks due to their technical, abstract, and problem solving natures, which often requires high-levels of skill and preparation. And this demand for workers with higher levels of skill and preparation is increasing relative to nonroutine skills [92].

According to the World Economic Forum, 36% of all jobs across all industries will require complex problem-solving skills by 2020, compared to 4% of jobs where basic physical abilities are a core requirement [300]. Similarly, MGI identifies that AI and automation will accelerate the shift from basic. skills[1] to higher skills[2]. For instance, demand for advanced IT and programming skills could grow as much as 90% between 2016 and 2030, whereas basic data input and processing skills could experience a decline of 23% over the same period [92].

This highlights the difficulties of AI, and computerisation more generally, to automate nonroutine tasks that are typically considered 'higher skilled' human tasks [87]. By comparison, routine tasks (both manual and cognitive) that are considered 'lower skilled' and highly repetitive, such as types of product assembly in manufacturing, are at greater risks of automation [151]. The proliferation of AI has the potential to accelerate this trend.

The preparation requirements for jobs are also increasing. And education and training systems have been slow to adapt to the increased skill requirements caused by technological changes [257]. This is in the context of an accelerating reskilling cycle, where workers are having to 'update' their skills at faster rates to keep apace with

[1] Basic cognitive skills include skills such as: basic literacy, numeracy, communication, basic data input and processing.

[2] Higher cognitive skills include skills such as: advanced literacy and writing, quantitative and statistical skills, critical thinking and decision making, project management, complex information processing and interpretation, and creativity.

technological changes [298]. While AI is not the sole contributor to these labour market changes, it has potential to accelerate and account for a significant proportion of these dynamics, particularly if AI fulfils projections as the next GPT [89].

These dynamics have created an effect referred to as 'skill-biased technical change' [164]. This is where new technologies increase the relative demand for high-skilled labour while reducing the demand for low-skilled labour. Acemoglu and Autor highlight this divergence by showing the growing wage premium earned by individuals with higher levels of education from 1963 to 2008 in the US [7]. The authors show that the wage premiums of 'college educated graduates' have significantly grown relative to 'high-school dropouts' and 'high-school graduates'. This increase in relative wages for educated labour also occurred in a period where the supply of educated labour increased [86]. The combination of higher pay and growing supply of high skilled labour clearly indicates the increase of relative demand for higher skilled labour.

The labour demands of AI threaten to exacerbate this trend; lower skilled labour and routine tasks are more susceptible to automation, and AI technologies complement higher skilled labour and drive up their relative demand [154]. These dynamics could have serious implications for rising income inequality, which will be discussed next.

2.3.8 AI and Inequality

The World Economic Forum identified income and wealth inequality as among the biggest global risks [340]. While technological progress is one of several factors that affect inequality, the rapid developments of AI rightfully raise questions whether the benefits will be equitably distributed. Democracies depend on the benefits of growth to be reasonably shared to ensure social cohesion and equality of opportunity. This is just as relevant for Australia as it is for all national economies.

2.3.8.1 Current state of Inequality in Australia

The state of economic inequality in Australia is complicated and subject to interpretation. This reflects the complexities of inequality measurement and evaluation. From a national average perspective, income and wealth inequality have remained relatively constant over the past few decades [283]. There was a general trend of increased income inequality until the Global Financial Crisis, which thereafter stabilised and is again at risk following the COVID-19 crisis [344]. According to the Gini coefficient, a common measure for income inequality, Australia's income inequality ranks slightly above average compared to other advanced OECD economies [259].

Beneath the averages, however, are signs of rising economic inequality. For instance, the top 1% and 10% of income earners have commanded consistently higher shares of national income in Australia since 1980 [344]. This is also in the context of increasing wage growth for higher earners, with higher levels of education, compared to middle and lower-income earners, with lower levels of education [256].

Layering these changes on top of entrenched disadvantage adds to the complexity. A sizeable part of Australia's population remains left behind. As of 2018, 13% of Australians from 0-17 years below the poverty line [260]. These levels of inequality are more likely to affect particular groups of the population, such as Indigenous Australians and people with a disability [283]. The location of where a child grows up also has a causal effect on their adult income outcomes in Australia, favouring urban dwellers [283]. Regardless of whether one interprets inequality in Australia to be a huge issue, it is subject to change. If AI is to fulfill its projected economic impacts as the next GPT, then it will certainly have profound structural effects on the Australian economy. How these economic benefits are distributed will influence economic and social inequality outcomes.

2.3.8.2 Relationship between technology and economic inequality

The historical relationship between transformational technologies and inequality depends on the length of time examined. The broad sweep of technological progress has improved inequality by lifting productivity, expanding the demand for labour, and increasing income, wealth, and quality of life [237]. This progress, however, was not immediate and often required more than 50 years for economies to adjust and widely diffuse its applications [194]. Therefore, the short-run disruptions of transformational technologies have caused profound structural changes to labour markets and economic activity [310]. These initial decades have typically required significant labour transitions and have contributed to widening short-run inequalities [84].

In comparison to other GPTs, the impacts of AI are likely to be a continuation of this 'short term pain for long-term gain' trend. That is, the adoption and diffusion of new technologies expand markets which affects the supply, demand, and mobility of resources and labour [117]. These new technologies improve productivity for industries, populations, and individuals to varying extents. This skews the distribution of benefits to those with the skills to make productive use of the new technologies [234]. As a result, wage premiums are earned by those with the skills that complement these technological changes, which can cause or exacerbate economic inequality. Additionally, as the share of income shifts from labour to capital, tax collection also becomes more difficult for governments, which can strain public revenues [2]. This process of 'creative destruction' by technological progress has been a relative constant in history of human innovation since the Agricultural Revolution [297]. In the long-run, people adapt, the overall demand for labour is reinforced, and inequality lessens.

2.3.8.3 Risks of AI to inequality

The rise of AI has led to cries of widespread labour automation from all corners of industry. Predictions have varied from 9 to 47% of current occupations in advanced economies [22, 151], like Australia [136]. This has led to headline grabbing claims of technological unemployment and human obsolescence. Yet, as discussed earlier in Section 2.3.6, occupations consist of many tasks that are difficult to automate [226]. In addition, these headline predictions neglect the new tasks and occupations that will arise courtesy of AI. While AI-enabled automation is expected to have a significant impact on the Australian economy, this must be balanced with the economic and social benefits that AI will provide.

The focus, therefore, should not centre on whether AI will destroy jobs; societies have regularly adapted to industrial and labour transformations from previous GPTs [80]. Rather, the focus should reorient towards the types of new skills and jobs demanded by AI, how to equip people with these skills to efficiently transitions between jobs, and the implications on inequality if the Australian labour market is slow, or fails, to transition to meet these new economic demands.

Growing skills divide. Risks to inequality arise when the 'race between education and technology' heavily favours technology [164]. That is, the rate of technological change outpaces the speed at which people can develop the new skills demanded by technology. Inequalities of wages then emerge as the demand for labour skills that complement new technologies increase and attract wage premiums. For example, skills that are non-routine and cognitive, such as abstract thinking in Machine Learning development, benefit from advances in AI due to strong complementarities between routine and cognitive tasks [42]. This raises the productivity and demand for workers with complementary skills to technology, thus driving up their wages.

The problem is that these skills, and subsequent wage premiums, disproportionately favour the highly educated (as discussed in Section 2.3.6). Australians with a bachelor degree in 2015 earned, on average, a 40% wage premium compared to someone with a high school diploma; a master's degree or doctorate earned a 79% premium [256]. And since 2001, these premiums are up from 20% and 55%, respectively [251]. There are risks that the growth of AI will further exacerbate this trend. This is problematic for inequality because jobs demanded by AI will likely require higher levels of skills and different mindsets, which could be difficult, or impossible, to develop for many workers.

AI technologies are also more likely to replace, rather than augment, routine tasks [155]. Such tasks are disproportionately found in low to middle-skilled occupations with lower levels of education [151]. Therefore, these low and middle-skilled jobs, which are already missing out on the wage premium, are also more exposed to AI-enabled labour automation and shifts in skill demands [49].

Employment polarisation. Skill shortages result in unfavourable economic and social outcomes [152, 284]. If only a small and shrinking proportion of the labour market can fulfil these high-skilled jobs, it places downward pressure on everyone else. More people enter the pools of lower skilled work and wages decrease as more people slide down the skill curve. Meanwhile, wages in higher-skilled labour disproportionately rise.

The effect of growing displacement of low and medium-skilled labour is referred to as 'employment polarisation' [14]. This is where labour supply becomes concentrated at either ends of the skill spectrum, which can obstruct upward social mobility. If employment polarisation worsens, there are fewer opportunities for people to climb the skill ladder, as the middle-skilled rung is weakened or shifted.

This process of turnover, accelerated by AI-enabled automation, could lead to sustained periods of underemployment or unemployment. Not all workers will have the training, skills, or safety-nets to successfully transition into the new jobs created by AI [57]. It can also result in the widening of income inequality [206]. As income from wages represents the majority of income for most households, and this dependence on wage income increases for poorer households [279], then widespread labour automation threatens economic security. Unless managed well, the transition effects of labour displacement caused by AI could have serious implications on economic inequality, and reinforce existing inequalities in Australia.

The costs of redistribution. If AI increasingly acts as a substitute for human labour, either because AI replaces workers or workers fail to transition into newly demanded occupations, then the economic benefits created by AI could be unequally distributed. Therefore, Korinek and Stiglitz argue that the primary economic challenge with the growth of AI is one of income distribution [206]. This refers to how income is distributed among members of a society given the economic changes brought about through AI.

Since markets rarely function perfectly in the real world [314], redistribution is generally needed. This helps to ensure that technological progress generates improvements for everyone, not just for innovators and investors endowed with capital. Income redistribution methods, either through direct transfers or funding social policies, such as labour market programmes, have played important roles in securing peace and stability of developed nations [312].

Balancing the incentives to innovate with the broader economic impacts on society is challenging and highly political. From an economic perspective, it depends on the costs of income redistribution. This refers to the transfer of income and wealth from some individuals to others through mechanisms such as taxation, welfare, and fiscal policy.

When the costs of redistribution are low in well-functioning markets, it is possible for the progress made by AI to benefit everyone [206]. There can be political agreement that the benefits of AI will be felt universally, so progress is desirable. If redistribution is too costly, however, and markets are not functioning equitably, then it may be difficult or impossible to compensate those adversely affected [206]. This could result in jobs being displaced without subsistence compensation, rising unemployment and underemployment, and widening economic inequality. Those at risk of experiencing such unwanted effects could rationally oppose AI developments, compromising social and economic stability. Therefore, market structures and institutions play an integral role in the costs of redistribution through social, fiscal, and monetary policies [313].

Korinek and Stiglitz further argue that there are two main ways for AI progress to affect the distribution of income and resources, and thus inequality: [206]

1. *Surplus earned by innovators*: As innovators and investors fund the development of innovations like AI, they rightfully expect to earn returns if they yield economic benefits. If, however, markets are not adequately contestable, due to reasons such as imbalanced intellectual property rights, then inefficiencies can emerge through disproportionate market power. This results in the surplus or net income earned by innovators and investors to be greater (sometimes significantly greater) than the costs of the innovation [205]. These returns earned by innovators and investors may not proportionately reflect the social returns to the innovation. As a result, redistribution can become costly because of monopolistic or oligopolistic markets, which can contribute to economic inequality.

2. *Effects on others*: Innovation can also lead to changes of income distribution in

an economy that directly impacts people who are not involved in the process of innovation. For example, an abrupt change in labour demand through automation. Economists refer to these redistributional impacts as 'externalities' from innovation. There are two main types: (1) *financial impacts*, such as wages, property, and costs of goods and services; (2) *non-financial impacts*, such as social and environmental effects, including drops in morale from unemployment or underemployment and pollution alleviation.

Social implications of rising inequality. If economic inequality were to acutely rise due to the effects of AI, then not only could the growth of AI be inhibited, but also the risks of social fragmentation could increase. In scenarios where workers are displaced by AI, and they do not receive adequate transition support or subsistence compensation, those affected could rationally oppose AI developments [206]. If a large part of the population do not economically benefit from the growth of AI, it is rational that they would defend their economic position [148]. This rejection of modernity could compromise social and economic development. As a result, AI is less likely to be adopted and diffused throughout the economy, which hampers economic growth, and also fuels political discontent because its benefits are being inequitably distributed. This is not a recipe for peace and democratic order. Rising inequality threatens social stability, which is highlighted by the positive correlation between rising income inequality and crime rates, both within and between countries [140].

It is predicted that the Australian workers who are more likely to be adversely affected by AI are also more likely to experience current levels of inequality, due to factors such as lower levels of education, socioeconomic disadvantage, and skill sets at greater risk of labour automation [117, 136]. It is therefore critical for the benefits of AI to be distributed equitably. Unless this is achieved, AI threatens to perpetuate these entrenched disadvantages, which is harmful to Australia economically and socially.

Mitigating the rise of AI-accelerated inequality. Public institutions play a central role in determining market structures that affect economic distribution [312]. This role is difficult, demanding a precarious balance between encouraging innovation on one hand, and ensuring its benefits are shared equitably on the other. Calo identifies a range of public policy mechanisms to help equitably distribute the benefits of AI in societies: [94]

- *Taxation and redistribution*: Applying effective tax and redistribution systems to ensure that the surpluses earned by innovators and investors help to support those inadvertently impacted by AI. This is typically performed through progressive taxation and transfers, which provides workers with subsistence compensation during periods of employment transition.

- *Infrastructure*: Effective digital infrastructures that help to diffuse AI equitably, such as 5G mobile networks and standards that foster open-data sharing. Infrastructures, such as Internet connectivity and access to digital devices, provide the backbone for the diffusion of AI. In a country as large and dispersed as Australia, ensuring equitable access to these critical infrastructures affects the extent of benefits that AI provides, particularly for rural and remote populations.

- *Antitrust policies*: Regulating anti-competitive behaviours by ensuring that companies do not stifle market competition and exhibit rent-seeking behaviours that adversely affect innovation and the consumer.

- *Intellectual property rights*: Creating incentives for companies to innovate by granting patents, but also ensuring that these exclusive rights do not unfairly block barriers to market entry.

- *Education and training*: Investing in the development of high-demand skills for youth, such as Science, Technology, Engineering, and Mathematics.

- *Minimum wage*: Helping to ensure that no one who works full time is in poverty.

- *Public research*: In parallel with effective antitrust policies, public research can help reduce the scope for monopolies that capture large portions of innovation returns. Innovations that are funded by public expenditure can be owned by the State and achieve market returns that contribute to public revenue, such as the CSIRO WiFi patent [116].

- *Labour Market programmes*: Policy measures that support displaced workers to transition into re-employment efficiently, such as targeted training programmes, social support services, and subsistence compensation. This is the area of focus for this research.

2.3.9 AI Adoption

The labour market impacts of AI depend on the adoption rates of AI technologies by firms. If firms are slow or fail to adopt AI, then its effects are naturally restricted. Therefore, the risks of AI accelerated labour automation will only be realised if these technologies are adopted by firms, absorbed in workflows, and broadly diffused. Otherwise, they're just isolated use cases.

This consideration, however, is often ignored. Much of the recent research on the economic impacts of AI assume broad adoption and diffusion. For example, the prominent study by Frey and Osborne estimated that 47% of occupations face a near-term risk of automation from AI [151]. These results were based on the assessments of a small panel of Machine Learning experts who were asked to identify which of 70 jobs were 'completely automatable' in 2013. However, these forecasts rely on some questionable assumptions. Chief among them is that firms will quickly and efficiently adopt AI for commercial use. This should not be taken as a given. As Bessen et al. [67, 69] point out that, just because new technologies have commercial applications does not mean that they will be adopted and diffused in a timely manner. Therefore, understanding the factors that influence the adoption and diffusion of AI in firms is important. It enables more accurate forecasting and better planning for policymakers, businesses, and civil society.

Explanatory variables for AI adoption and diffusion. Research on the factors that affect firms' decisions to adopt digital technologies is well established [188, 198, 291]. Researchers have closely examined the adoption dynamics of innovations such as personal computers [320], the Internet [19], and social media [277]. AI builds upon these digital technologies. The factors that influence the adoption of AI by firms differ by degree but not by kind. The literature suggests eight major factors influencing AI adoption rates at the firm-level:

(1) *Competition*: McKinsey Global Institute found that the extent of rivalry within markets has the largest effect on AI adoption [93]. This is consistent with game theory [238], where the marginal propensity to adopt AI depends on the proportion of rivals that have already decided to adopt. Assuming the new technology becomes broadly diffused, then early adopters typically enjoy disproportionate rewards. However, as more firms adopt, the marginal incentive to adopt diminishes as the technology provides less competitive advantages. Therefore, laggard firms are punished with shrinking market shares [20]. These competitive forces drive adoption rates as firms jostle to assert a competitive edge and advance market share [222]. However, adoption decisions are made with imperfect information as it can be difficult to know what actions competitors are taking. Competition, therefore, can drive rapid periods of adoption growth.

(2) *Firm characteristics*: The size, income level, and industry of firms have all been

shown to affect the rate that a new technology is adopted [171]. For example, larger firms, by headcount and income, typically adopt digital technologies earlier and at faster rates than smaller firms. Also, firms in Financial Services and ICT industries tend to adopt digital technologies earlier and at faster rates than firms in Agricultural and Construction industries [27]. Similarly, the AI adoption indicator we propose suggests material differences between industry categories, with highest levels of adoption in Financial & Insurance Services firms and lowest in the Agriculture Industry.

(3) *Labour force skill capabilities*: Emerging technologies, such as AI, often require specific skills [67]. The availability of workers with these skills can influence the extent of adoption and diffusion [56]. The ability to access such labour competencies, however, varies between firms, industries, and economies. The implementation of AI requires strong technical competencies. These competencies are unevenly distributed between firms, industries, and economies [69]. Therefore, the more firms are able to access relevant skilled labour, the greater the likelihood that firms will adopt AI.

(4) *Digital Maturity*: Previous research has shown that the adoption of new digital technologies often depends on the adoption of previous digital technologies [21]. For instance, broadband infrastructure supports the adoption of more sophisticated digital applications. This relationship also appears to hold for AI. According to McKinsey Global Institute, firms that have adopted and absorbed cloud infrastructure and 'web 2.0 technologies', such as mobile technologies and Customer Relationship Management (CRM) systems, are more likely to adopt AI technologies [93].

(5) *Expected return on AI investments*: Firms' perceptions of the value that a new technology can create also influences adoption rates [18]. Similarly, firms that are positive about the business use cases of AI are more likely to adopt earlier and faster [93]. Conversely, firms that are uncertain about AI use cases are slower or less likely to adopt, which delays aggregate adoption rates.

(6) *AI complements*: As with other General Purpose Technologies, the more complementary technologies are developed and implemented, the faster AI will be adopted by firms [89]. That is, the more a firm invests in one type of AI, the more likely it will invest in another. For example, a retailer that implements robotic process automation to retrieve stock is more likely to adopt computer vision to identify inventory items than a retailer that has not adopted any AI technologies. Capital investment deepens as AI is increasingly absorbed in workflows.

(7) *Regulatory effects*: Regulatory effects can be important to consider when comparing the adoption rates across economies [188, 275]. For example, it is plausible that the more

stringent data protection regulations in Europe could delay AI adoption in European firms compared to US firms in the short-run.

(8) *Standardisation and usability*: As the use of emerging technologies are standardised across firms and industries, the ease of use for these technologies naturally improves, which has been shown to accelerate adoption and diffusion rates [68]. While AI models are still 'narrow' in the sense that they tend to be highly specific to a particular task and require non-routine customisation (as in hyper-parameter tuning or feature data engineering), AI usability has improved over the past decade. For example, individuals are able to implement high-performing machine learning models using their own data with little or no knowledge of scripting languages (see [166]). As the use of AI technologies become standardised and usability improves, it is likely that this will increase adoption rates.

While other variables could affect rates of AI adoption, the eight factors listed above are likely to account for a significant proportion of firm-level AI adoption decisions.

2.4 Positioning this research

The papers presented in this thesis showcase data-driven methods for analysing skill shortages, job transitions, and AI adoption in Australia. This builds upon the previous works aforementioned in this literature review and provides novel approaches that use big data from a range of labour market sources. The growth and sophistication of data science and machine learning techniques, coupled with the the increased availability of large and detailed data sources on labour markets, has enabled unique and cross-disciplinary avenues for analysing skill shortages, job transitions, and new technology adoption. Therefore, the papers in this thesis approach these established problems of labour economics with a set of data science and machine learning techniques that are applied to detailed datasets from job ads, employment statistics, household surveys, and other occupational data. As per the requirements of a 'Thesis by Publication', the following papers both standalone in their own right (having already been published or currently under-review) and also serve the broader research themes of this thesis. That is, to analyse the changing labour market dynamics in Australia from 2012 to 2020 with a specific focus on skill shortages, job transitions, and the demand for skills that enable AI adoption at the firm-level.

Chapter 3

Paper 1 - Adaptively selecting occupations to detect skill shortages from online job ads

Preamble

This paper identifies a range of variables and creates data-driven measures to detect skill shortages from real-time job ads data. It then applies these methods to a targeted case study of Data Science and Analytics (DSA) occupations in Australia from 2012-2019. DSA jobs are selected because they are the primary occupation class responsible for implementing AI technologies within firms and they are often referred to as occupations being 'in shortage' [70, 213, 225]. To select DSA occupations, we put forward a method to adaptively select occupations from their underlying skills. This allows us to identify occupations whose underlying skills are evolving and increasingly requiring AI and DSA-related skills. The significance of this research is threefold: first, we develop a data-driven methodology to identify occupations of interest from their underlying skill demands, as opposed to arbitrary selection from standardised and static occupation titles; second, we identify and construct five key variables for detecting skill shortages from real-time job ads data; and third, we apply the aforementioned methods to analyse DSA occupations in Australia as the primary occupation group for implementing AI technologies. The research presented in this paper directly relates to the broader thesis themes of the skill shortages and AI technology adoption. This paper was published in the *2019 IEEE International Conference on Big Data (Big Data)* proceedings [120].

Abstract

Labour demand and skill shortages have historically been difficult to assess given the high costs of conducting representative surveys and the inherent delays of these indicators. This is particularly consequential for fast developing skills and occupations, such as those relating to Data Science and Analytics (DSA). This paper develops a data-driven solution to detecting skill shortages from online job advertisements (ads) data. We first propose a method to generate sets of highly similar skills based on a set of seed skills from job ads. This provides researchers with a novel method to adaptively select occupations based on granular skills data. Next, we apply this adaptive skills similarity technique to a dataset of over 6.7 million Australian job ads in order to identify occupations with the highest proportions of DSA skills. This uncovers 306,577 DSA job ads across 23 occupational classes from 2012-2019. Finally, we propose five variables for detecting skill shortages from online job ads: (1) posting frequency; (2) salary levels; (3) education requirements; (4) experience demands; and (5) job ad posting predictability. This contributes further evidence to the goal of detecting skills shortages in real-time. In conducting this analysis, we also find strong evidence of skills shortages in Australia for highly technical DSA skills and occupations. These results provide insights to Data Science researchers, educators, and policy-makers from other advanced economies about the types of skills that should be cultivated to meet growing DSA labour demands in the future.

3.1 Introduction

The Internet has become the primary channel for disseminating information in many areas of society. This is the case for job advertisements (ads), where approximately 60% of Australian job ads are posted online [126]. At aggregate levels, online job ads can provide valuable indicators of relative labour demands. Rather than relying solely on lagging indicators from labour market surveys, online job ads data can reveal shifting labour demands as they occur. This can provide policy-makers, researchers, and businesses with additional data points to assess the health and dynamics of labour markets.

Real-time labour demand data is essential for Data Science and Analytics (DSA) occupations because of how rapidly DSA skills are evolving and diffusing into other occupational classes. In this research, DSA skills refer to the use of scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data, which can be used to make data-driven decisions and actions [131].

DSA skills are multi-disciplinary, adopting methods from fields such as statistics, mathematics, and computer science. A distinction can also be made between skills, knowledge, abilities, and occupations. 'Skills' are the proficiencies developed through training and/or experience [263]; 'knowledge' is the theoretical and/or practical understanding of an area; 'ability' is the competency to achieve a task [157]; and 'occupations' are the amalgamation of skills, knowledge, and abilities that are used by an individual to perform a set of tasks that are required by their vocation. For simplicity, throughout this paper the term 'skill' will include 'knowledge' and 'ability'.

There are several challenges when analysing the labour demands of occupations and assessing the extent of skills shortages. The first challenge concerns accurately identifying occupations based on their evolving skill demands. Occupations are organised into standardised hierarchical classifications, which vary across national jurisdictions. Most often, these are static, rarely updated classifications, which fail to capture the changing skill demands, or to detect the creation of new occupations. For instance, 'Data Scientists', 'Data Engineers' and 'Data Analysts' do not exist in the Australian and New Zealand Standard Classification of Occupations (ANZSCO); rather, they are all grouped as 'ICT Business Analysts'. Furthermore, even when occupations are analysed based on their skill frequencies [157], biases emerge from the difference in their relative frequency. For example, 'Communication Skills' occur in around one-quarter of all job ads used in this work. However, just because some skills are common does not mean that they are more or less important than other skills that are also required in an individual job. This leads to two related questions: (1) **how to adaptively identify relevant skills from labour market data while minimising biases that emerge from ad hoc aggregations?** And (2) **how to identify relevant occupations based on this generated set of skills?**

The second challenge is detecting evidence of skills shortages from (near) real-time data. Skill shortages are mostly measured via labour market surveys [244]. This involves surveying employers about their abilities to access workers who possess the skills their firms demand. A major shortcoming of this approach is that surveys are difficult to scale, and that they are rarely conducted on statistically valid samples [125]. Another significant issue is that labour market surveys are lagging indicators, i.e. the publication of results can be many months after the data was collected. Lastly, due to scaling limitations, prominent labour market surveys on skills shortages (or mismatches) fail to measure all standardised occupations [263]. Therefore, the questions are **can we detect evidence of skill shortages from real-time labour market data? If so, what are**

the key variables for assessing skills shortages from such data?

This paper addresses the above challenges using a large dataset of over 6.7 million Australian online job ads spanning between 2012-01-01 and 2019-02-28, which has been generously provided by Burning Glass Technologies[1] (BGT). The data has been collected via web scraping and systematically processed into structured formats. The dataset consists of detailed information on individual job ads, such as location, salary, employer, educational requirements, experience demands, and more. The skill requirements have also been extracted (totalling $> 11,000$ unique skills) and each job ad is classified into its relevant occupational and industry classes.

To address the first challenge, we first adapt an established similarity measure originating from Trade Economics [185] to measure the pairwise similarity between unique skills in job ads. Next, we develop a novel data-driven method to generate sets of skills highly similar to a set of seed skills. Finally, we uncover the relevant occupations for which at least 15% of all skills required in their associated ads are from the target set of skills. We apply this method to uncover the set of DSA skills and DSA occupations, starting from a seed set of common DSA skills.

We address the second challenge by identifying five key variables from online job ads data which are critical for detecting skill shortages in real-time: (1) job ad posting frequency; (2) median salary levels; (3) educational requirements; (4) experience demands; and (5) job posting predictability. We then analyse the DSA occupations according to each of these five variables and find compelling evidence for how these features are predictive of skill shortages.

The main contributions of this work include:

- We develop a **data-driven methodology to construct skills sets** for specific occupational areas, and to select occupations based on granular skills-level data;
- We identify **five key variables for detecting skill shortages from online job ads data**;
- We apply the aforementioned methods to a unique dataset of online job ads to **analyse the changing labour demands of DSA skills and occupations** in the advanced economy of Australia. We also **construct and share the list of top DSA skills** generated from this dataset.

[1]BGT is a leading vendor of online job ads data. `https://www.burning-glass.com/`

3.2 Related Work & Limitations

Job ads data as a proxy for labour demand. During 2001-2003, Lee [211] gathered job ads data from the websites of Fortune 500 companies in order to analyse the skill requirements of 'Systems Analysts'. Lee was able to determine that these positions demanded their candidates to have 'all-round' capabilities, beyond just technical skills. More recently, Gardiner et al. [157] procured 1,216 job ads with 'Big Data' in the job title from the `indeed.com` API. The authors then conducted content analyses to investigate how 'Big Data' skills have manifested in labour demand. Their research reiterated that employers are demanding technical skills in conjunction with 'softer' skills, such as communication and team-work.

DSA skill shortages. While the capacity to collect, store, and process information may have sharply risen, it is argued that these advances have far outstripped present capacities to analyse and make productive use of such information [184]. Claims of DSA skill shortages are being made in labour markets around the world [70, 213, 225], including in Australia (see Supplementary Materials in Section 3.8 for more details). Most similar to this research, however, are two studies conducted using BGT data to assess DSA labour demands. The first was an industry research collaboration between BGT, IBM, and the Business-Higher Education Forum in the US [227]. The research found that in 2017 DSA jobs earned a wage premium of more than US$8,700 and DSA job postings were projected to grow 15% by 2020, which is significantly higher than average. In another study commissioned by the The Royal Society UK [70], BGT data were analysed for DSA jobs in the UK. The results also showed high levels of demand for DSA skills, particularly 'technically rigorous' DSA skills.

Limitations of using online job ads data. It is argued that job ads data are an incomplete representation of labour demand. Some employers continue to use traditional forms of advertising for vacancies, such as newspaper classifieds, their own hiring platforms, or recruitment agency procurement. Job ads data also over-represent occupations with higher-skill requirements and higher wages, colloquially referred to as 'white collar' jobs [101] (see Supplementary Materials in Section 3.8 for more details).

Occupational classifications. There are significant shortcomings in official occupational standards. Official occupational classifications, like ANZSCO, are often static taxonomies and are rarely updated. We therefore use the BGT occupational classifications because of its adaptive taxonomies that update with changing labour demands. For example, a job ad title of 'Senior Data Scientist' is classified as a 'Data Scientist' in the

BGT occupational classification but is classified as an 'ICT Business & Systems Analyst' by ANZSCO. For more details, please review the Supplementary Materials (Section 3.8).

3.3 Skill similarity and sets of related skills

Intuition. Skills provide the means for workers to perform labour tasks in order to fulfill their occupational demands. Therefore, the assortment of skills required for a job, and their pairwise interconnections uniquely identify occupations. In this section, we propose a methodology to capture the 'similarity' between skill-pairs that co-occur in job ads. Intuitively, two skills are similar when the two are related and complementary, i.e. the skills-pair supports each other. For example, 'Python' and 'TensorFlow' have a high similarity score because together they enable higher productivity for the worker, and because the difficulty to acquire either skill when one is already possessed by a worker is relatively low.

The Revealed Comparative Advantage of a skill. We develop a data-driven methodology to measure the pairwise similarity between pairs of skills that co-occur in job ads. One difficulty we encounter is that some skills are ubiquitous, occurring across many job ads and occupations. We address this issue by adapting the methodology proposed by Alabdulkareem et al. [14] to maximise the amount of skill-level information obtained from each job ad, while minimising the biases introduced by over-expressed skills in job ads. We use the *Revealed Comparative Advantage* (RCA) to measure the relevance of a skill s for a particular job ad j, computed as:

$$RCA(j,s) = \frac{x(j,s) / \sum_{s' \in \mathscr{S}} x(j,s')}{\sum_{j' \in J} x(j',s) / \sum_{j' \in \mathscr{J}, s' \in \mathscr{S}} x(j',s')}$$

where $x(j,s) = 1$ when the skill s is required for job j, and $x(j,s) = 0$ otherwise; $\mathscr{S}$ is the set of all distinct skills, and $\mathscr{J}$ is the set of all job ads in our dataset. $RCA(j,s) \in \left[0, \sum_{j' \in J, s' \in S} x(j',s')\right], \forall j,s$, and the higher $RCA(j,s)$ the higher is the comparative advantage that s is considered to have for j. Visibly, $RCA(j,s)$ decreases when the skill s is more ubiquitous (i.e. when $\sum_{j' \in J} x(j',s)$ increases), or when many other skills are required for the job j (i.e. when $\sum_{s' \in S} x(j,s')$ increases).

RCA provides a method to measure the importance of a skill in a job ad, relative to the total share of demand for that skill in all job ads. It has been applied across a range of disciplines, such as trade economics [185] [330], identifying key industries in nations [307], and detecting the labour polarisation of workplace skills [14].

Measure skill similarity. The next step is measuring the complementarity of skill-pairs that co-occur in job ads. First we introduce the 'effective use of skills' $e(j,s)$ defined as $e(j,s) = 1$ when $RCA(j,s) > 1$ and $e(j,s) = 0$ otherwise. Finally, we introduce the skill complementarity (denoted θ) as the minimum of the conditional probabilities of a skills-pair being effectively used within the same job ad. Skills s and s' are considered as highly complementary if they tend to commonly co-occur within individual job ads, for whatever reason. Formally:

$$\theta(s,s') = \frac{\sum_{j' \in J} e(j,s).e(j,s')}{max\left(\sum_{j' \in J} e(j,s), \sum_{j' \in J} e(j,s')\right)}$$

Note that $\theta(s,s') \in [0,1]$, a larger value indicates that s and s' are more similar, and it reaches the maximum value when s and s' always co-occur (i.e. they never appear separately).

Top DSA skills. We use the θ function to create a list of DSA skills. First, we qualitatively select 5 common DSA skills as seed inputs: *'Artificial Intelligence', 'Big Data', 'Data Mining', 'Data Science'*, and *'Machine Learning'*. Next, for each of these 5 DSA skills, we calculate the top 300 skills with the highest similarity scores. Finally, we merge the five lists, we calculate the average similarity scores for each unique skill, and rank in descending order. This results in a ranked list of 589 skills, which we qualitatively assess and decide keep the top 150 skills. While some skills outside of the top 150 could be considered DSA skills, it was at this point that the relevance to DSA skills began to deteriorate and merge into other domains. For example, skills such as *'Design Thinking', 'Front-end Development'*, and *'Atlassian JIRA'* – which are technical, but not DSA specific – were just outside of the top 150 skills.

The purpose of this top DSA skills list is to capture DSA labour trends rather than represent a complete taxonomy of DSA skills. The list of top 150 DSA skills can viewed in the Supplementary Materials (Section 3.8).

3.4 DSA occupations and categories

Compute the skill intensity. In this section, we present an adaptative technique to uncover DSA occupations from job data. First, we compute η the *'DSA skill intensity'* for each standardised BGT occupation, defined as percentage of DSA skills relative to the total skill count for the job ads related to an occupation o. Formally:

$$\eta(o,\mathcal{D}) = \frac{\sum_{j \in \mathcal{O}, s \in \mathcal{D}} x(j,s)}{\sum_{j \in \mathcal{O}, s' \in S} x(j,s')}$$

where $\mathcal{D}$ is the set of DSA skills, and $\mathcal{O}$ is the set of job ads associated with the occupation o.

Select the top DSA occupations. We qualitatively assessed the occupational list ordered by η, and decided to establish a cutoff at $\eta > 15\%$. The rationale for this threshold level was that occupations just below this cutoff are questionably considered DSA occupations – take for example, *'Web Developer'* and *'UI / UX Designer / Developer'*. Occupations just above this threshold appeared more consistent with the definition of DSA skills given in the *Introduction*. Moreover, the occupations with a DSA skill intensity level just above the 15% threshold represented occupations where the authors considered DSA skills to likely become more prevalent. For example, the demands for DSA skills are expected to increase for 'Economists' due to the growing amounts of economic data that are being made available [138]. Therefore, this list represents occupations where DSA skills are already important, or have reached a minimum threshold of DSA skill intensity and where DSA skills are likely to become more important for the occupation.

Table 3.1 shows the 23 occupational classes that satisfy these DSA threshold requirements. Occupations are categorised to compare labour dynamics within the DSA occupational set. The occupational categories are adapted from previous BGT research completed in the US [227] and UK [70]. Here, BGT grouped DSA occupations into categories based on skill similarities and sorted categories according to 'analytical rigour' of their skill sets [70, 227]. We have applied their categorical framework here because (1) we are using the equivalent BGT dataset for the Australian labour market and (2) many of the DSA occupations uncovered in this research are also present and categorised in their studies. Fig. 3.1 illustrates the categorical framework, giving a brief definition of each category and places them on a comparative scale of 'analytical rigour'.

3.5 Detecting Skill Shortages from Job Ads

In this section, we propose five labour demand variables for detecting skill shortages from job ads data. These include: (1) job ad posting frequency growth; (2) median salary levels; (3) educational requirements; (4) experience demands; and (5) job posting predictability.

Table 3.1: Selected DSA Occupations and their job ad counts.

DSA Category	DSA Occupation	#Ads
Data Scientists and Advanced Analysts	Biostatistician	270
	Computer Scientist	38
	Data Engineer	71
	Data Scientist	2,388
	Economist	2,127
	Financial Quantitative Analyst	947
	Mathematician	105
	Physicist	423
	Robotics Engineer	18
	Statistician	2,535
Data Analyst	Business Intelligence Architect / Developer	3,166
	Data / Data Mining Analyst	34,520
Data Systems Developers	Computer Programmer	16,311
	Computer Systems Engineer / Architect	73,437
	Data Warehousing Specialist	964
	Database Administrator	17,937
	Database Architect	7,489
	Mobile Applications Developer	4,357
	Software Developer / Engineer	113,247
Functional Analysts	Business Intelligence Analyst	23,547
	Fraud Examiner / Analyst	653
	Security / Defense Intelligence Analyst	482
	Test Technician	1,592
TOTALS	**23 DSA Occupations**	**306,577**

Analytical Rigor

Category	Functional Role
Data Scientists and Advanced Analysts	Create sophisticated analytical models used to build new datasets and derive new insights from data
Data Analysts	Leverage data analysis and modeling techniques to solve problems and glean insight across functional domains
Data Systems Developers	Design, build and maintain organisational data and analytical infrastructure
Functional Analysts	Utilise data and analytical models to inform specific functions and business decisions

Figure 3.1: Defining DSA Categories adopted from Markow et al and Blake [70, 227]

We argue that these variables taken together provide explanatory insight for identifying skill shortages of occupations.

3.5.1 Variables for detecting skill shortages

This research has found evidence of DSA skill shortages for the 'Data Scientists and Advanced Analysts' ('Data Scientists', henceforth) and 'Data Analysts' categories. A combination of factors have led to these conclusions.

Job ads posting frequency. Both categories have experienced high relative growth in terms of posting frequencies (shown in Fig. 3.2(a)). High posting frequency growth can be indicative of increasing employer demands for workers that possess specific occupational skills [247]. Both 'Data Scientists' and 'Data Analysts' have averaged higher than average year-on-year growth rates (28% and 13%, respectively) than the other DSA categories and the market average (10%) (see Fig. 3.2(b)).

Salaries. 'Data Scientists' and 'Data Analysts' command high, and growing, wage premiums (Fig. 3.2(c)). High and growing wages indicate that employers are willing to pay a premium to attract workers with specific skills [95]. That is, when labour supply is constrained and labour demand increases, then wages should increase, as is the case for 'Data Scientists' and 'Data Analysts'.

Education levels. High relative educational requirements can constrain the supply of skilled labour by creating barriers to entry [95]. In Fig. 3.2(d), this is especially evident for 'Data Scientists', where the years of education required by employers is significantly higher than average and other categories.

Experience demands. The minimum years of experience demanded by employers can vary according to the accessibility of skilled labour. If employers have difficulty hiring the labour they demand, then they may reduce their experience-level requirements as part of their recruitment efforts [183]. As Fig. 3.2(e) shows, this is again the case for 'Data Scientists' and to a lesser extent 'Data Analysts', where experience levels have remained relatively low. For 'Data Scientists', the minimum experience requirements have decreased by almost one year since 2012 and sit just above the market average. For 'Data Analysts', the average years of minimum experience have been below the market average since after 2016.

Job ad posting predictability. In the next subsection we report on predictability of job postings and argue that predictability of job ad posting frequency should be considered as an explanatory variable for detecting skill shortages. To anticipate the findings of that section, we note the difficulties of predicting occupations (and skills)

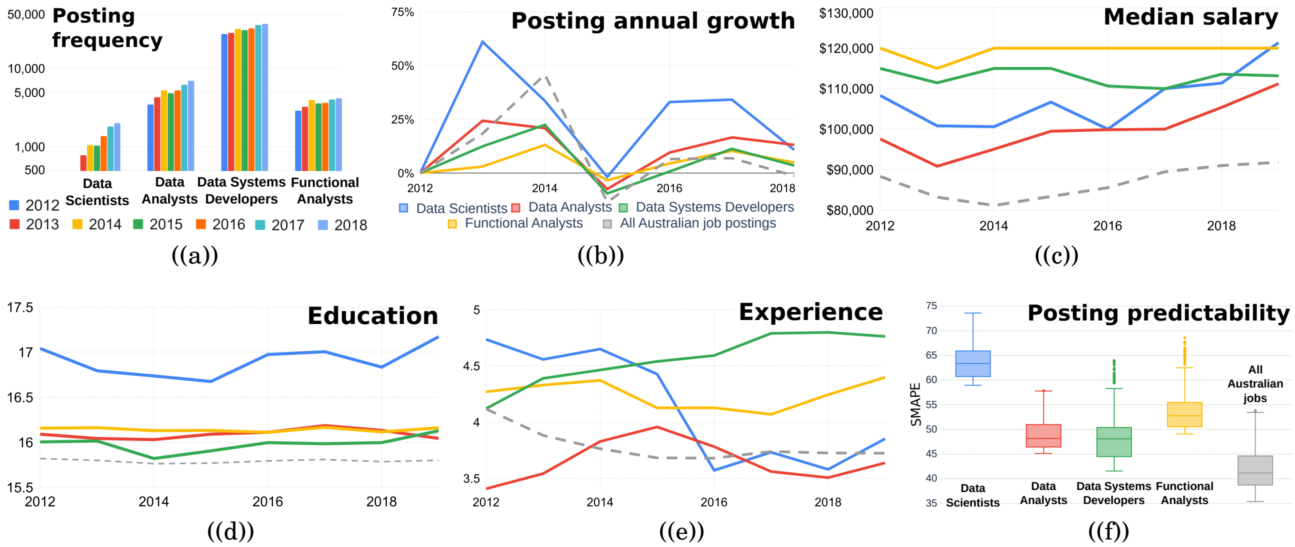


Figure 3.2: Labour demand variables for detecting skill shortages from job ads data: posting frequency (a) and its annual growth (b); median salary (Australian \$) (c); education level (years of formal education) (d); experience (years) (e) and job ad posting predictability in terms of SMAPE error scores (f).

that have high-growth in terms of job ad postings. As seen in Fig. 3.2(f), the forecast predictions for 'Data Scientists' job ads perform relatively poorly compared to the lower growth categories. We contend that this is due to the rapidly changing labour dynamics of 'Data Scientists' and that this lack of predictability tends to highlight the patterns of high-growth occupations, reflecting another measure of rising labour demands. In the next section (Section 3.5.2) we detail how we quantify the predictability variable.

Taken collectively, these factors form a strong case that the Australian labour market has been experiencing a shortage of 'Data Scientists' and 'Data Analysts'. These variables form a framework of features to detect skill shortages from job ads.

3.5.2 Predict job ad posting

Forecast ad postings. In this section, we propose a 'predictability' feature by building a time series model to predict job ad posting frequencies for each of the categories. We use the Prophet time series forecasting tool developed by Facebook Research [319]. Prophet is an auto-regressive tool that fits non-linear time series trends with the effects from daily, weekly, and yearly seasonality, and also holidays. The three main model components are represented in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{3.1}$$

where $g(t)$ refers to the trend function that models non-periodic changes over time; $s(t)$ represents periodic changes, such as seasonality; $h(t)$ denotes holiday effects; and ϵ_t is the error term and represents all other idiosyncratic changes. For more details on Prophet and its hyper-parameter choices, please refer to the Supplementary Materials (Section 3.8).

Prediction error measure. Using Eq. (3.1), one can run forward time and forecast job ad posting frequency. We measure the accuracy of the forecast using the Symmetric Mean Absolute Percentage Error (SMAPE) [220, 301]. SMAPE is formally defined as:

$$SMAPE(A_t, F_t) = \frac{200}{T} \sum_{t=1}^{T} \frac{|F_t - A_t|}{(|A_t| + |F_t|)}$$

where A_t denotes the actual value of jobs posted on day t, and F_t is the predicted value of job ads on day t. SMAPE ranges from 0 to 200, with 0 indicating a perfect prediction and 200 the largest possible error. When actual and predicted values are both 0, we define SMAPE to be 0. We selected SMAPE as an alternative to MAPE because it is (1) scale-independent and (2) can handle actual or predicted zero values. For a discussion on alternate error metrics, please consult the Supplementary Materials (Section 3.8).

Evaluation protocol.

The forecasts made using Prophet are deterministic (i.e. given the same input, we will obtains the same output). We evaluate the uncertainty of predicted future job ad volumes using a 'sliding window' approach. As shown in Fig. 3.3, we use a constant number of training days (here $1,186$ days) to train the model, and we test the forecasting performance on the next 365 days. We shift both the training and the testing periods right by one day, and we repeat the process. We iterate this process 365 times, denoted in Fig. 3.3 using *Train start* for the starting point of the train period, *Test start* for the starting point of the test period, and using *Window start* for the starting point of the unused period. Consequently, we train and test the model 365 times, and we obtain 365 SMAPE performance values, which are presented aggregated as a boxplot in Fig. 3.2(f). The advantage of this approach is that it provides a distribution of SMAPE scores across a range of testing periods, which allows for a more robust evaluation of the modelling performance.

3.6 Discussion

Job ad posting trends ($g(t)$ in Eq. (3.1)) have grown for all DSA categories since 2012. This is shown in Fig. 3.4, which isolates $g(t)$ for each category to highlight the non-

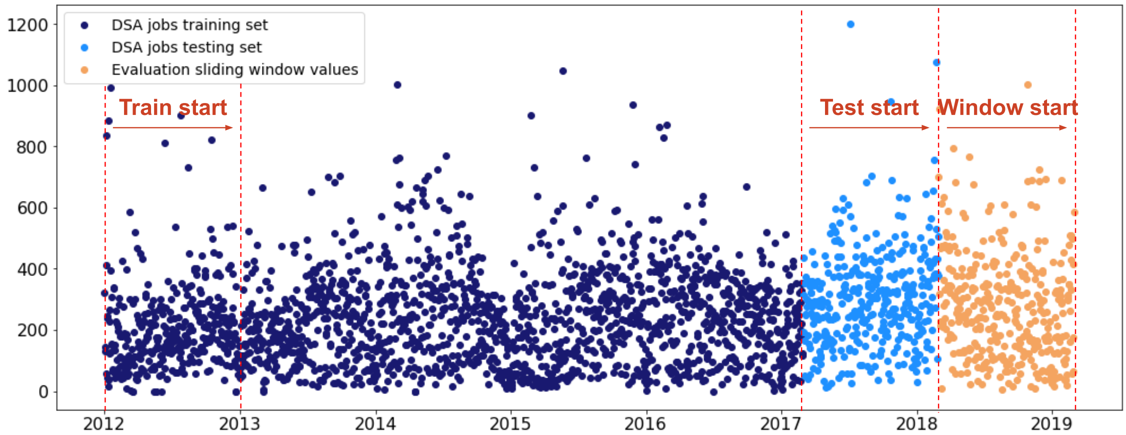


Figure 3.3: Sliding window setup for evaluating job ads forecasting performance.

periodic changes of daily job ad posts. Here, Fig. 3.4 shows that the more technically rigorous categories of 'Data Scientists' and 'Data Analysts' have experienced the highest growth trends. There are three distinct change point periods observed in Fig. 3.4. Firstly, from January 2012 to April 2014, where the frequency of all job ads are growing. Over this period, only 'Data Scientists' grew at a faster rate than the total market for 'All Australian Job' Ads (using the simple growth formula). This period can perhaps be explained by (1) the higher levels of job openings being posted online earlier in the dataset and (2) the early stages of DSA skills demanded by occupations, particularly for the more technically rigorous occupations.

The second period, from approximately May 2014 to November 2017, was generally one of slowing growth for online job ads. A possible explanation for this period is Australia's increasing underemployment rate [29]. Underemployment rose relatively steeply from just above 7% in 2014, diverging from a lowering unemployment rate, before reaching a peak just below 9% around the beginning of 2017. Underemployment then began to slightly decrease until the end of 2018. The sharp rise in underemployment could be indicative of employers being less willing or able to hire due to softening labour market conditions, which would presumably affect the frequency of job ad postings. While the more analytically rigorous categories of 'Data Scientists' and 'Data Analysts' also experienced slowing growth, they both grew at higher rates relative to other categories. The fact that these categories maintained strong upward trends, despite dampening labour market forces, highlights the high levels of labour demand for these occupational categories.

The final period from October 2017 until February 2019 (the end of this dataset), was

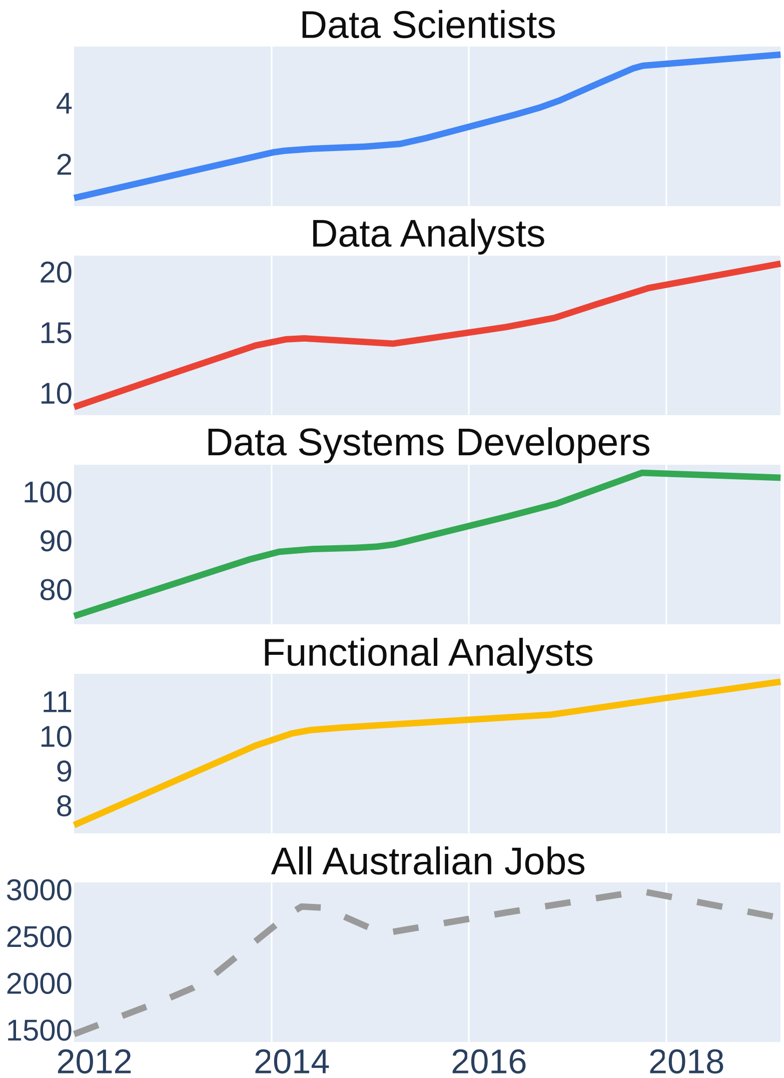


Figure 3.4: Trend lines of daily online job ad postings

generally one of stagnation or slight growth. Again, 'Data Scientists' and 'Data Analysts' continued upward trajectories, albeit at slower growth rates than previous periods. All DSA categories had higher trend growth rates than 'All Australian Job Postings' during this period. This final change point period highlights some possible conclusions. Firstly, the frequency of online job ads have potentially reached a saturation point. This means that the maximum proportion of job postings captured via online aggregators might have reached its upper limits. If this is the case, then any posting frequency growth for specific occupational classes above the total market rate could indicate high (or relatively high) labour demand. From this perspective, all DSA jobs continue to experience higher labour demands relative to all Australian job ads postings in the dataset since 2014.

The strong relative growth of 'Data Scientists' and 'Data Analysts' also provides insight. One interpretation is that Australian firms and employers have started to increasingly adopt AI technologies. A recent report by McKinsey & Co suggests that this is the case [318]. The accelerating rate of AI adoption requires highly skilled labour to make productive use of these technologies. These are the same analytically rigorous skills that are demanded from 'Data Scientists' and 'Data Analysts'. As a result, some portion of this growing labour demand for DSA skills, particularly the highly technical DSA skills, could be explained by accelerating AI adoption by Australian firms. Another related perspective is that Australian firms have increasing access to data with potentially meaningful insights. Therefore, workers with DSA skills that are able to productively use and draw insights from such data would logically be in high demand.

3.7 Conclusions and Future Research

In this research, we firstly developed a data-driven methodology to construct an adaptive set of skills highly similar to a set of seed skills. We then applied this method to identify the DSA skills set and DSA occupations, organising these occupations into common DSA categories. Secondly, we proposed five variables from online job ads data which are critical for the real-time detection of skill shortages. We then analysed the DSA categories according to each of these five variables. Here, we find strong evidence for how these features are collectively predictive of skill shortages. From this analysis, we find evidence that Australia is experiencing skills shortages for 'Data Scientists' and 'Data Analysts' occupations. A combination of indicators points to these conclusions. Firstly, both categories have experienced high relative growth in terms of **job ad posting frequencies**. Secondly, both categories command high, and growing, **wage premiums**. Thirdly, both categories demand higher than average **education requirements**, which constrains the supply of skilled labour pursuing these vocations. This is especially the case for 'Data Scientists'. Fourthly, the average minimum years of **experience** required by employers for these categories are low. For 'Data Scientists', the minimum experience requirements have decreased by almost one year since 2012 and sit just above the market average. For 'Data Analysts', the average years of minimum experience have been below the market average since 2017. Lastly, these occupational categories are relatively **difficult to predict**, especially for occupations in the 'Data Scientists' category. Taken collectively, these factors form a strong case that the Australian labour market has been experiencing a shortage of 'Data Scientists' and 'Data Analysts'.

Limitations and future work. A limitation of this work is that it only consists of labour demand data, and estimates labour supply via the proxy of the five proposed variables. Future work will corroborate these findings according to official labour shortage lists published by governments (i.e. a labour supply 'ground truth'). This could be achieved by developing a multivariate logistic classifier where the five proposed variables are used as features to predict whether an occupation is experiencing shortage. Conducting equivalent analyses on other markets and occupational groups could also provide insights into the predictive performance of these explanatory variables.

3.8 Supplementary Materials

This section accompanies the published paper, entitled *Adaptively selecting occupations to detect skill shortages from online job ads*. The information in this document complements the paper, and it is presented here for completeness reasons. It is not required for understanding the main paper, nor for reproducing the results.

3.8.1 Australia's looming DSA Shortfall

The Australian Computer Society (ACS), Australia's peak body group for Information and Communication Technologies (ICTs), forecasts that Australia will need almost 100,000 additional ICT professionals just to keep up with demand by 2023 [123]. Approximately half of these ICT professionals will require highly technical or digital management skills. However, domestic completions of ICT degrees were just 5,502 in 2016 [36]. This current level of labour supply is insufficient to meet the future demands for ICT professionals generally, and DSA occupations specifically.

3.8.2 Limitations of Online Job Ads Data

The biases discussed in Section 3.2 are present in the dataset used for this research. For example, 52.8% of Australian job ads in the dataset were classified as 'Professionals' or 'Managers' in 2018 (39.5% and 13.3%, respectively), according to the official Australian and New Zealand Standard Classification for Occupations (ANZSCO). These are typically 'white collar' occupations. In comparison, employment data from the Australian Bureau of Statistics (ABS) indicates that 'Professionals' and 'Managers' collectively represent just 36.2% of employment in Australia (23.7% and 12.5%, respectively) [31]. The traditionally 'blue collar' workers from categories such as 'Machinery Operators and Drivers' and 'Labourers' appear to be underrepresented in the BGT dataset.

Similarly, the 2018 average salary range for all online job ads in Australia was AUD$89,028 - $98,904. This is higher than the average full-time wage in Australia, which was $83,408 in November 2018 [28]. Therefore, as online job ads fail to cover the universe of employment vacancies, they should be interpreted as trends rather than 'ground truth' for labour demand. However, these biases do not impede this research too significantly, as a major component of this research is comparing different classes of DSA jobs, which are all considered in the 'Professionals' or 'Managers' classes.

3.8.3 Challenges with Classifying Occupations

A general challenge with classifying job ads is that job titles are not uniform. A 'Senior Machine Learning Engineer' and a 'Deep Learning Specialist' have different job titles but may require the same skills. Therefore, they should be measured in the same occupational class. An issue with ANZSCO, however, is that classifications are rarely updated; the last review was in 2013. So, emerging skills are not always properly captured or can be missed entirely leading to inaccurate classifications. So, the two example occupations above may be classified into different occupational classes despite having consistent skill requirements. Misclassified occupations can distort true representations of labour markets. Additionally, emerging skills, such as many DSA skills, complicate static and rarely updated occupational classifications.

These challenges have led BGT to develop their own taxonomies of labour skills and occupational classifications. BGT currently maintain a dictionary of over 11 thousand job skills. When processing job ads, BGT extract the skill requirements for each job. Typically there are multiple skill requirements for a unique job. For example, a 'Data Science' job could consist of the following skills: 'Python', 'SQL', 'Data Warehousing', 'Communication Skills', and 'Team Work / Collaboration'. These job skills build the foundation of BGT's adaptive occupational classification system.

3.8.4 DSA Skill Demands

Comparing relative DSA skill demands involved counting the frequency of each DSA skill that occurs in unique DSA job ads. As seen in Fig. 3.5, Structured Query Language ('SQL') has consistently been the DSA skill in the highest demand.

The Compound Annual Growth Rate (CAGR) was calculated for each DSA skill based on their first available posting in a DSA job, which is shown in Table 3.2. 'Blockchain' has unequivocally been the fastest growing DSA skill. However, this growth has been over a short period of time, with its first recording in 2016. The other fastest growing DSA skills have generally been analytical tools used either to manage 'Big Data' or 'Artificial Intelligence' (AI) related techniques. For instance, 'Apache' Spark', 'Apache Kafka', and [Apache] 'PIG' are all open source software tools used to assist with 'Big Data' management and processing. Additionally, skills such as 'TensorFlow', 'Deep Learning', and 'Random Forests' are all skills that generally pertain to AI.

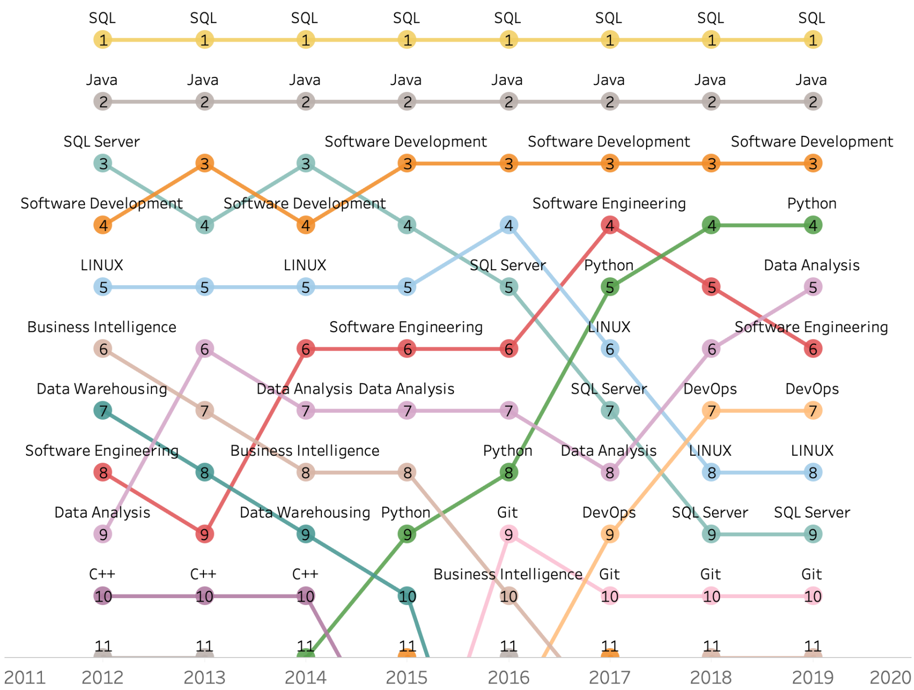


Figure 3.5: Top 10 DSA Skills for each year from 2012-2019

3.8.5 Time Series Forecasting with Prophet

Time series analysis provides a set of techniques to draw inferences from a sequence of observations stored in time order [81]. The development of accurate time series models can offer insights into the principal components that have affected historical growth trajectory patterns. They also facilitate a means for making predictions into the future.

This paper applies a relatively new and high-performing time series forecasting tool developed by Facebook, called Prophet [319]. The forecasting tool is applied to Australian online job ads data to uncover growth trends of DSA jobs.

In 2017, Facebook Research released Prophet as an open source forecasting procedure implemented in the Python and R programming languages. When benchmarked against ARIMA, ETS (error, trend, seasonality) forecasting, seasonal naive forecasting, and the TBATS model, Prophet forecasts had significantly lower Mean Absolute Percentage Errors (MAPE) [319].

The default hyperparameters of Prophet were applied for this analysis. This included an uncertainty interval of 80%, the automatic detection of trend change points, and the estimations of seasonality using a partial Fourier sum. For seasonality, Prophet uses a Fourier order of 3 for weekly seasonality and 10 for yearly seasonality. Experimentation steps were conducted by specifying a custom holidays dataframe, adjusting smoothing

Table 3.2: Top DSA Skills Growth

Rank	DSA Skill	CAGR
1	Blockchain	616%
2	TensorFlow	283%
3	Apache Spark	271%
4	Deep Learning	201%
5	Apache Kafka	188%
6	Internet of Things (IoT)	182%
7	Microsoft Power BI	175%
8	Data Lakes / Reservoirs	169%
9	Qlik	157%
10	Random Forests	151%
11	Apache Hive	145%
12	PIG	136%
13	Pipeline (Computing)	134%
14	Supervised Learning (Machine Learning)	131%
15	Boosting (Machine Learning)	129%
16	Alteryx	128%
17	Sqoop	119%
18	Apache Flume	109%
19	DevOps	107%
20	Unsupervised Learning	102%

parameters, and fitting the model with a multiplicative seasonality setting. However, all of these specifications led to a slight deterioration of performance metrics. Therefore, the default hyperparameters were restored, which the authors state *"are appropriate for most forecasting problems"*[319].

3.8.6 Evaluating performance

The Prophet library includes a method for calculating a range of evaluation metrics.[2] However, these metrics are not ideal for measuring prediction performance of online job ads for two reasons.

Firstly, analyses in this paper are comparing DSA categories with different scales of job posting frequencies. Therefore, most metrics calculated by Prophet's diagnostics method, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root

[2]The method is called `cross_validation`. For more information, see: https://facebook.github.io/prophet/docs/diagnostics.html

Mean Square Error (RMSE), are not suitable for comparisons because such measurements are scale-dependant [122].

Secondly, an appropriate performance metric for this dataset must not be distorted by zero values. This is important for job posts, where some DSA categories recorded zero daily postings, particularly earlier in the dataset. Subsequently, this rules out the last meaningful performance metric calculated by Prophet's diagnostics, namely MAPE. As the dataset contains zero values for posting frequencies, MAPE values can be infinite as it involves division by zero.

Therefore, accommodating for these two criterion points, the selected prediction performance metric is the Symmetric Mean Absolute Percentage Error (SMAPE). SMAPE is an alternative to MAPE that is (1) scale-independent and (2) can handle actual or predicted zero values. SMAPE, first proposed by Armstrong [301] and then by Makridakis [220], is defined by Section 3.5.2.

3.8.7 DSA Skills List

Table 3.3 shows the selected DSA skills, i.e. the top 150 skills selected using the methodology described in Section 3.3.

Table 3.3: Selected 150 Data Science and Analytics skills.

Rank	Skill	Theta	Rank	Skill	Theta
1	Machine Learning	0.375157109	76	Computer Vision	0.016534028
2	Data Science	0.339677644	77	Ruby	0.016521212
3	Big Data	0.281395532	78	Microsoft Sql Server Integration Services (SSIS)	0.016224833
4	Data Mining	0.275784695	79	PostgreSQL	0.015755516
5	Artificial Intelligence	0.268911214	80	Informatica	0.015750079
6	Apache Hadoop	0.160263705	81	Applied Statistics	0.014990736
7	R	0.120578077	82	SQL Server Reporting Services (SSRS)	0.01460998
8	Big Data Analytics	0.11683186	83	Data Management	0.014488424
9	Predictive Models	0.087256126	84	Data Lakes / Reservoirs	0.014444455
10	Scala	0.078168962	85	Metadata	0.014422194
11	Tableau	0.071103958	86	Quantitative Analysis	0.014245931
12	Apache Hive	0.068540161	87	Qlik	0.013849961
13	Python	0.067852169	88	ElasticSearch	0.013784912
14	SAS	0.058335431	89	Information Retrieval	0.013626625
15	NoSQL	0.054171879	90	Scalability Design	0.013495411
16	Teradata	0.053266061	91	Database Design	0.013409781
17	SPSS	0.052294251	92	Apache Flume	0.013268289
18	Natural Language Processing	0.051589073	93	Supervised Learning (Machine Learning)	0.013255296
19	MATLAB	0.049969987	94	Regression Algorithms	0.013068441
20	Data Visualisation	0.049141083	95	Model Building	0.012974866
21	Data Transformation	0.043785348	96	Visual Basic for Applications (VBA)	0.012941596
22	MapReduce	0.04200936	97	PERL Scripting Language	0.012885431
23	Data Modelling	0.041207512	98	Cognos Impromptu	0.012817815
24	Statistical Analysis	0.040950811	99	SAP BusinessObjects	0.012601388
25	Predictive Analytics	0.040725603	100	Oracle Business Intelligence Enterprise Edition (OBIEE)	0.012256767
26	Statistics	0.040600659	101	Prototyping	0.012183407
27	Deep Learning	0.040097617	102	Node.js	0.012089477
28	Internet of Things (IoT)	0.038865379	103	Experimental Design	0.012083924
29	PIG	0.038346523	104	MySQL	0.012051979
30	Extraction Transformation and Loading (ETL)	0.037375468	105	Classification Algorithms	0.01192503
31	Data Architecture	0.037357392	106	Logistic Regression	0.011923395
32	Data Warehousing	0.037120923	107	Relational DataBase Management System (RDBMS)	0.011907611
33	Microsoft Power BI	0.03691897	108	Statistical Methods	0.011798527
34	Apache Kafka	0.03478849	109	Splunk	0.0116979
35	Neural Networks	0.034594775	110	Sqoop	0.011619513
36	Data Engineering	0.033870742	111	GitHub	0.011606854
37	Econometrics	0.033635451	112	Unsupervised Learning	0.011432418
38	Data Integration	0.031413571	113	Apache Impala	0.011420459
39	Data Structures	0.029579863	114	Web Analytics	0.011406332
40	Decision Trees	0.029538939	115	Git	0.011202096
41	Business Intelligence	0.028968279	116	Amazon Web Services (AWS)	0.01118572
42	C++	0.028931884	117	Datastage	0.011123658
43	Pipeline (Computing)	0.027558689	118	Optimisation	0.011085172
44	Consumer Behaviour	0.0273288	119	Simulation	0.010785033
45	Hadoop Cloudera	0.027221747	120	LINUX	0.010773868
46	Data Quality	0.0264852	121	Software Development	0.010750719
47	Clustering	0.026032976	122	Continuous Integration (CI)	0.010688564
48	Apache Webserver	0.026020174	123	Business Intelligence Reporting	0.010349562
49	Qlikview	0.025944556	124	Agile Development	0.010225424
50	Cassandra	0.025060662	125	Solution Architecture	0.010225063
51	Consumer Research	0.024973131	126	AWS Elastic Compute Cloud (EC2)	0.010217691
52	Apache Spark	0.024017603	127	Microstrategy	0.010147521
53	AWS Redshift	0.023822744	128	Marketing Analytics	0.010006654
54	Data Manipulation	0.023299597	129	Bash	0.009937595
55	Cluster Analysis	0.022795077	130	Alteryx	0.009881429
56	Microsoft Azure	0.022690165	131	SQL Server	0.009830543
57	Experiments	0.022525239	132	Shell Scripting	0.009614866
58	Physics	0.021968001	133	Credit Risk	0.009534963
59	Software Engineering	0.020672929	134	Image Processing	0.009483378
60	Cloud Computing	0.020237968	135	Boosting (Machine Learning)	0.009409621
61	MongoDB	0.020228716	136	Platform as a Service (PaaS)	0.009390802
62	Consumer Segmentation	0.0202243	137	Transact-SQL	0.009342661
63	DevOps	0.020103595	138	Version Control	0.009182692
64	Relational Databases	0.01974885	139	Support Vector Machines (SVM)	0.009167358
65	Data Analysis	0.019621418	140	Data Warehouse Processing	0.00903522
66	Blockchain	0.019568638	141	Customer Acquisition	0.009029462
67	Data Governance	0.019300535	142	Linear Regression	0.008983594
68	SQL	0.019192807	143	Software Architecture	0.008952848
69	SQL Server Analysis Services (SSAS)	0.018858212	144	Google Analytics	0.008950648
70	Java	0.018541708	145	AWS Simple Storage Service (S3)	0.008939552
71	TensorFlow	0.018237584	146	Dimensional and Relational Modelling	0.008727614
72	Text Mining	0.017501842	147	Microsoft SQL	0.008714559
73	Random Forests	0.0173648	148	Functional Programming	0.008700033
74	Robotics	0.01663332	149	Scrum	0.008677026
75	Distributed Computing	0.01659359	150	Economics	0.008593447

CHAPTER 4

Paper 2 - Predicting Skill Shortages in Labor Markets: A Machine Learning Approach

Preamble

This paper develops high performing machine learning models to predict occupational skill shortages in Australia ahead of time. It does so by connecting related datasets from job ads, employment statistics, and occupational skill shortage classifications, and applying state-of-the-art machine learning techniques for tabular data. Further to the prediction tasks, we put forward a method to analyse the underlying skill demands of occupations in shortage, using 'Data Scientists' as an example. We also measure the feature importance from the prediction models to better understand which variables are most predictive of skill shortages. This research paper contributes a novel data-driven method to analyse and predict skill shortages with high levels of performance. It also advances the understanding of the major factors contributing to occupational skill shortages. This paper relates to the core theme of skill shortages in this thesis. The paper was published in the *The 4th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in Big Data* as part of the *2020 IEEE International Conference on Big Data* [121].

Abstract

Skill shortages are a drain on society. They hamper economic opportunities for individuals, slow growth for firms, and impede labor productivity in aggregate. Therefore, the ability to understand and predict skill shortages in advance is critical for policy-makers and educators to help alleviate their adverse effects. This research implements a high-performing Machine Learning approach to predict occupational skill shortages. In addition, we demonstrate methods to analyze the underlying skill demands of occupations in shortage and the most important features for predicting skill shortages. For this work, we compile a unique dataset of both Labor Demand and Labor Supply occupational data in Australia from 2012 to 2018. This includes data from 7.7 million job advertisements (ads) and 20 official labor force measures. We use these data as explanatory variables and leverage the XGBoost classifier to predict yearly skills shortage classifications for 132 standardized occupations. The models we construct achieve macro-F1 average performance scores of up to 83 per cent. Our results show that job ads data and employment statistics were the highest performing feature sets for predicting year-to-year skills shortage changes for occupations. We also find that features such as 'Hours Worked', years of 'Education', years of 'Experience', and median 'Salary' are highly important features for predicting occupational skill shortages. This research provides a robust data-driven approach for predicting and analyzing skill shortages, which can assist policy-makers, educators, and businesses to prepare for the future of work.

4.1 Introduction

In January 2019, Andrew Penn, the CEO of Telstra – Australia's largest Telecommunications company – announced that the company will be expanding its new 'Innovation and Capability Center' in Bangalore, India. This will create approximately 300 Network and Software Engineering jobs, with the potential for more [274]. Penn cited 'skill shortages' as the main reason for this outsourcing decision:

> "We need these capabilities now, but the fact is we cannot find in Australia enough of the skills that we need on the scale that we need them, such as software engineers. Why? There simply are not enough of them. The pipeline is too small." [103]

This coincides with Telstra announcing a goal net reduction of 8,000 jobs by 2022 (mainly in Australia), as the company seeks to automate labor tasks and simplify processes [106]. While an isolated example, the evolving labor demands of Telstra highlights both the opportunity costs of skill shortages and the precariousness of workers' security to automation and globalization. As a result of these claimed skill shortages, the Australian labor market will not enjoy the benefits afforded by 300 highly skilled jobs – benefits that materialize in greater economic activity, labor productivity, and economic competitiveness. This is not specific to just Telstra or Australia, skill shortages burden most labor markets to varying extents [85]. Their impacts limit employment opportunities for individuals, impede technology adoption and investment by firms, and hamper labor productivity in aggregate.

In this work, we focus on three open problems relating to skill shortages at the occupational level. The first problem relates to analyzing the underlying skills of occupations known or suspected to be in shortage. Skills enable workers to complete labor tasks that are required by jobs. Therefore, analyzing the demand and relative importance of skills within occupations provides granular insights into which skills should be developed and prioritized for occupations in shortage. This can help to inform policy-makers, educators, and individual job-seekers. However, most approaches to determining skill importance within an occupation have relied on *ad hoc* aggregations of job advertisements (ads) or resume data [157, 211]. While simple frequency counts can provide useful proxies for demand, such methods do not normalize for highly common skills and can therefore yield distorted views of skill importance within occupations. So, the question is **(1) can we determine which skills are most important for occupations in shortage while accounting for highly common skills?**

The second open problem is concerned with predicting occupational skill shortages. While the adverse effects of skill shortages have been well-documented [85, 175, 262], predicting skill shortages is difficult. Even more challenging is predicting temporal changes to the skill shortage status of an occupation. For example, accurately predicting whether an occupation will shift from being classified as *Not in Shortage* in one time period to *In Shortage* the next. These difficulties reflect the lack of consensus around which variables are most predictive of skill shortages and the limited available data classifying occupational shortages. The question is, therefore, **(2) can we leverage modern Data Science and Machine Learning techniques to predict occupational skill shortages?**

The third open problem relates to understanding which variables are most predictive of skill shortages. While many studies have examined the presence of skill shortages in labor markets [104, 175, 196], there remains a lack of understanding about which factors contribute most to occupational shortages. This leads to the final question: by building predictive models, **(3) can we uncover which variables are most important for predicting skill shortages at the occupational level?**

We address each of the above-stated questions by leveraging both labor demand and labor supply data. With regards to the first research question, we use a rich dataset of 7,697,568 job ads in Australia to analyze the underlying skill demands of 'Data Scientists', an occupation shown to be in shortage in Australia [120], the UK [70], and the US [227]. Here, we compare two different methods to assess the top temporal skill demands of 'Data Scientists'. We highlight the shortcomings of *ad hoc* skill counts and illustrate an alternative method that captures specialized and emerging skills within an occupation.

We address the second research question by constructing a supervised Machine Learning model framework to predict skills shortage classifications at the occupational level one year into the future. These binary classification models are built using eXtreme Gradient Boosting (XGBoost), a scalable Machine Learning system for tree boosting [108]. We incorporate labor demand and labor supply occupational data from Australia as input, which are organized and matched according to the official Australian occupational standards. On the labor demand side, we again use job ads data from the aforementioned dataset, spanning from 2012-01-01 to 2018-12-31. For the labor supply side, we use 'Detailed Labor Force' data from the Australian Bureau of Statistics over the same time period [31]. Lastly, the 'ground-truth' (or predictive variable) is taken from the longitudinal list of occupational shortages, recorded by the Australian Federal Department of Education, Skills and Employment [124]. These official skill shortage classifications directly inform national and state policies in the areas of education, training, employment and skilled immigration. Further detail on the data is discussed in Section 4.3.

Lastly, we address the third research question by extracting the feature importance data generated from the above prediction model. This sheds light on which variables are most important for predicting the skill shortage status of an occupation. Importantly, we find empirical evidence that 'Hours Worked', 'Education', 'Years of Experience', and 'Salary' are the most important features for predicting occupational skills shortages. This supports evidence from Labor Economics where workers in occupations experiencing skill shortages tend to have higher work intensity and longer work hours [180, 287]. Similarly, employers attempt to overcome skill shortages and meet labor demands by lowering education requirements, experience demands, and increase salary levels to attract a greater pool of candidates [85, 120, 180]. These variables prove to be predictive features (see Section 4.4.4).

The main contributions of this work are the following:

- We compare two methods to **analyze the underlying skill demands of occupations in shortage and detect emerging skills**, using 'Data Scientists' as the example;
- We implement a **data-driven modeling framework to predict temporal skill shortages of occupations**;
- Lastly, we **analyze the feature importance data from the prediction models to identify which variables are most predictive of skill shortages**.

4.2 Related Work & Limitations

We structure this discussion of the related work into two areas. First, in Section 4.2.1, we visit work dealing with measuring skill shortages. Second, in Section 4.2.2, we investigate the economic costs of skill shortages.

4.2.1 Measuring Labor Shortages

The broader problem. Skill shortages occur when the labor demand for specific skills exceed the supply of workers who possess those skills at a prevailing market wage [180, 196]. Skill shortages can be considered a subset of the broader problem of 'Skills Mismatch'. At the macro-level, skills mismatch refers to the disequilibrium of aggregate supply and demand of labor skills, usually with reference to a specific geographic unit [85]. Skill shortages are one scenario of skills mismatch and occur when the demand for specific skills exceed the available supply of workers at real wage rates. Conversely, 'Skill Surpluses' are caused by an excess of skill supply [284]. That is, there are more workers who possess specific skills that the labor market demands on aggregate. Therefore, skill shortages are usually calculated as a component of measuring skill mismatches.

For a discussion on the factors that cause skill shortages, please refer to the Supplementary Materials (see Section 4.7).

Measures using surveys. Skill shortages are typically measured at the firm-level through the use of surveys to examine the extent of unfilled and hard-to-fill vacancies [230]. A shortcoming of this approach is that skill shortages can be overstated and such surveys are often unrepresentative. For instance, employers may claim an occupation to be *In Shortage* but the underlying cause could be their own inability to offer a sufficient wage-level, attractive working conditions, or a desirable location. These micro-level factors can distort the presence of genuine skill shortages, where employers extrapolate their firm-specific challenges as macro-level issues [104] (see Supplementary Materials in Section 4.7 for further details).

Use of indirect measures. To differentiate between perceived and genuine skill shortages, other studies have complemented survey results with indirect measures, such as wage growth, employment growth, vacancy rates, and work intensity. The rationale underlying these approaches is that occupations experiencing skill shortages are typically characterised by wage premiums, greater employment growth, growing vacancy rates, and higher work hours and levels of overtime [85]. The OECD implemented such indirect measures in concert with employer surveys to construct a series of indicators and composite indexes on skills for employment, including skill shortages [257]. The 'World Indicators of Skills for Employment' (WISE) database calculates an occupational indicator of skill shortages based on wage growth, employment growth, and growth in the hours worked [253]. Next, this indicator is transformed into a composite skill index that uses the O*NET database [326] to map occupations into groups of skills and tasks. This allows for international comparability between OECD countries for skills challenges and performances, including the extent of skill shortages.

Other approaches have used indicators from job ads data to assess skill shortages. Dawson et al. [120] analyzed a large temporal dataset of online job ads to detect skill shortages of Data Science and Analytics occupations in Australia. The authors use a range of indicators to evaluate the presence and extent of skill shortages, such as posting frequency, salary levels, educational requirements, and experience demands. They contend that occupations experiencing high posting growth appear volatile and their posting frequencies are difficult to predict. Given that high and growing posting frequency is often used as a proxy for high labor demand for occupations, the authors argue that high error metrics, combined with the other indicators, can help detect skill shortages. In this work, we use the labor demand features features proposed in Dawson et al. [120] to build a skill shortage classifier. For completeness reasons, we describe these features in the Supplementary Materials (Section 4.7).

The current work. The present work takes a data-driven machine learning approach to measure and predict skill shortages. We leverage a set of recently proposed labor demand features extracted from job ads data [120], together with official labor supply features to build a machine learning model that classifies whether an occupation is in shortage. In addition, we analyze the relative importance of these features.

4.2.2 Economic Costs of Skill Shortages

The costs of skill shortages can be significant and manifest at both micro and macro-levels of economies. They affect individuals, firms, and aggregate markets.

Individual-level. Skill shortages can negatively affect earnings and reduce development opportunities for workers. Markets experiencing skill shortages can force individuals to accept less desirable and insecure work. In 2011, Quintini [284] analyzed household survey data from the European Community Household Panel to investigate the effects of qualification mismatch on earning. Quintini found that 'over-qualified' individuals earn approximately 3% less than individuals with the same occupations but who have been appropriately matched. The presence of skill shortages exacerbates the inefficient allocation of labor, which can negatively affect the earnings and employment opportunities for individuals.

Firm-level. Several studies have examined the implications of skill shortages on firm-level productivity and all concluded that skill shortages negatively impact firm-level productivity [63, 147, 176, 317]. In a study using the Australian Business Longitudinal Database, Healy et al. [180] found that most Australian firms respond to skill shortages through longer working hours and higher wages for occupations experiencing in shortage. Significantly, we found that the 'Hours Worked' and 'Salary' levels were among the most important features for predicting skill shortages, seen in Section 4.4.4. However, there is evidence to suggest that such skill shortages are usually short-lived. Further research analyzed the existence of skill shortages in German firms and concluded that while their effects can be acute, they are typically a temporary and short-term phenomena [62].

Macroeconomic-level. Lastly, the economic costs of skill shortages accumulate to macroeconomic effects. Frogner [152] uses data from the Employers Skill Survey to identify the negative impacts of skill shortages on productivity, Gross Domestic Product, employment levels, and wage earnings. From the perspective of private investment, Nickell et al. [245] calculates that a 10% increase in firms reporting skill shortages decreases private investment by 10% and Research & Development investment by 4%. The inefficient allocation of resources caused by skill shortages therefore hampers productivity, which can compromise macroeconomic growth.

The current work proposes a method to predict in advance skill shortages and better understand their contributing factors. These methods and results could in turn be used by policy-makers, educators, and companies to prepare for and alleviate the negative impacts of skill shortages.

4.3 Data and Methods

In this section, we first detail the data sources and the constructed labor demand and labor supply features (Section 4.3.1). We then outline two methods to assess skill importance for occupations classified as in shortage (Section 4.3.2. Last, we detail the prediction model setup and evaluation (Section 4.3.3).

4.3.1 Data sources and constructed features

In this work, we employ both labor demand and labor supply data as explanatory variables (features, henceforth) to predict occupational skill shortages. The dataset we construct relates to occupations in Australia during the period 2012-2018. Due to space constraints, the table summarizing all the constructed features is shown in the Supplementary Materials (Section 4.7).

Labor demand features. For labor demand, we have used job ads data, which was generously provided by Burning Glass Technologies[1] (BGT). The data has been collected via web scraping and systematically processed into structured formats. The dataset consists of detailed information on individual job ads, such as location, salary, employer, educational requirements, experience demands, and more. Each job ad is also categorized into its relevant occupational classification. We build upon the results of Dawson et al. [120] and we incorporate a range of the engineered job ads indicators that the authors found predictive of labor shortages, as discussed in Section 4.2.1.

While data from BGT integrates multiple online sources and arguably represents the most comprehensive repository of job ads data, it is argued that online job ads are an incomplete representation of labor demand [101], for two reasons. First, some employers continue to use traditional forms of advertising for vacancies, such as newspaper classifieds, their own hiring platforms, or recruitment agency procurement. Second, job ads data also over-represent occupations with higher-skill requirements and higher wages, colloquially referred to as 'white collar' jobs [101]. These are limitations of the current work, discussed in Section 4.6.

[1]BGT is a leading vendor of online job ads data: `https://www.burning-glass.com/`

Labor supply features. The labor supply data used for this research comes from the 'Quarterly Detailed Labor Force' statistics by the Australian Bureau of Statistics (ABS) [31]. This consists of statistics on employment levels, unemployment, underemployment, hours worked and others. As the labor supply statistics are measured quarterly, the yearly average for each feature was calculated to match the skills shortage target variable, which is measured in yearly periods (presented next).

Skill shortages ground-truth. The ground-truth comes from the 'Historical List of Skill Shortages in Australia', measured by the Australian Federal Department of Education, Skills and Employment (DESE, henceforth) [124]. For over three decades, the DESE has conducted ongoing skills shortage research in Australia. Their research aims to identify shortages for skilled occupations where long lead times for training means that such shortages cannot be addressed immediately. The DESE tracks 132 occupations nationally, and they also provide more detailed analyses on select occupations at the State and Territory levels. To assess skill shortages, the DESE survey employers every year, called the 'Survey of Employers who have Recently Advertised' (SERA). The SERA collects both qualitative data from employers and recruitment professionals, and quantifiable data on employers' recruitment experiences [38]. The output of this DESE activity is that, for every year, each of the 132 tracked occupations is classified as *In Shortage* or *Not In Shortage* at the national-level. The results of these classifications have direct implications for education, training, employment and migration policies.

There are, however, five important limitations of the DESE's methodology for measuring skill shortages. First, the DESE acknowledge that the survey is not a statistically valid sample of Australia's labor market. Second, there are inherent limitations of determining skill shortages from surveying employers, as discussed in Section 4.2.1. Nonetheless, the ABS evaluated the methodology and found that it was "appropriate for its purpose" [38]. **To our knowledge, this dataset is the most reliable source of occupational skill shortages that is publicly available in Australia.** Third, the surveyed occupations in this research are biased towards the occupational classes of 'Technicians and Trades' workers and 'Professionals'. Forth, the dataset is imbalanced with a greater number of occupations classified as *Not in Shortage*. Fifth and finally, there are inherent limitations that emerge from analyzing jobs using standardized occupational taxonomies. Specifically, official occupational classifications are usually static taxonomies that are rarely updated and slow to adapt to changing labor dynamics. This research uses the official Australian and New Zealand Standard Classification of Occupations (ANZSCO) [25]. While other more adaptive taxonomies exist, ANZSCO remains the official taxonomy and is the measurement standard used for all data in this research.

4.3.2 Quantify skill importance for occupations

Here, we detail two approaches for determining relative levels of skill importance for an occupation known or suspected to be in shortage. We exemplify both methods in Section 4.4.2 using job ads classified as 'Data Scientists' from 2015-2019 in Australia, as this occupation has be shown to be in shortage during this period [120, 123]. Analyzing the underlying skills of occupations in shortage is important as it provides granular details on which skills should be targeted to help alleviate occupational shortages. This assists policy-makers, educators, and job-seekers to prioritize the development of specific skills to help meet evolving labor demands.

Posting frequency as a proxy for demand. The proxy most widely used in literature [70, 101, 227] for skill importance is skill frequency – i.e. count how many times a skill appears in the job ads associated with a given occupation during a predetermined period of time. While skill frequency can provide some indication of labor demand (i.e. higher skill counts being indicative of higher demand), it fails to normalize for skills that are demanded by all or most jobs. This does not necessarily reveal which skills are more or less important to a given occupation, as some skills generalize across all occupations at high frequencies (for e.g. 'Communication Skills' and 'Teamwork'). This leads to an alternate method for assessing skill importance within occupations.

Normalized skill importance. Here, we use an established measure called 'Revealed Comparative Advantage' (RCA) that has been applied across a range of disciplines, such as trade economics [185, 330], identifying key industries in nations [307], and detecting the labor polarization of workplace skills [14]. RCA measures the importance of a skill in a job ad, relative to the total share of demand for that skill in all job ads. Formally, the RCA for skill s and the job ad j is:

$$RCA(j,s) = \frac{x(j,s) / \sum_{s' \in \mathscr{S}} x(j,s')}{\sum_{j' \in J} x(j',s) / \sum_{j' \in \mathscr{J}, s' \in \mathscr{S}} x(j',s')}$$

where $x(j,s) = 1$ when the skill s is required for job j, and $x(j,s) = 0$ otherwise; $\mathscr{S}$ is the set of all distinct skills, and $\mathscr{J}$ is the set of all job ads in our dataset. $RCA(j,s) \in \left[0, \sum_{j' \in J, s' \in S} x(j',s')\right], \forall j,s$, and the higher $RCA(j,s)$ the higher is the comparative advantage (or importance) that s is considered to have for j. Visibly, $RCA(j,s)$ decreases when the skill s is more common (i.e. when $\sum_{j' \in J} x(j', s)$ increases), or when many other skills are required for the job j (i.e. when $\sum_{s' \in S} x(j, s')$ increases). Therefore, RCA adjusts for the biases that emerge from high-occurring skills across all jobs, while maximizing the skill-level information within individual jobs.

We compute skill importance weights at the occupational level $W_{s,o}$ – i.e. how important is a particular skill s in the occupation o for year t – as the mean RCA for skill s in job ads pertaining to occupation o (denoted as J_o):

$$W_{s,o} = \frac{1}{|J_o|} \sum_{j \in J_o, j \in t} RCA(j, s)$$

As a last step, we sort the skills by $W_{s,o}$ in descending order, filtering out extremely rare skills that occur less than five times during a year. This returns a list of top skills that can be interpreted as the most important to occupation o for year t, adjusted for high-occurring skills. As is seen in Section 4.4.2, the resulting skills list from this method yields newly emerging and more specific skills than that of posting frequency.

4.3.3 Predictive Setup for Skill Shortages

Choosing a classification model. In this work, we predict skill shortages by employing XGBoost [108] – an off-the-shelf classification algorithm. XGBoost is an implementation of gradient boosted tree algorithms. XGBoost has achieved state-of-the-art results on many standard classification benchmarks and is a well established Machine Learning framework [270]. As an overview, these are Machine Learning techniques that produce prediction models in the form of an ensemble of weak prediction models (here decision trees), by optimizing a differentiable loss function [107]. We chose to use XGBoost because it is the currently the state-of-the-art in both classification and regression tasks for medium sized amounts of data (i.e. where neural networks cannot be fully deployed). It also features several advantages that we leverage in our regression task: it automatically handles missing data values, and supports parallelization of tree construction.

Accounting for the temporal inertia of shortage classifications. Skill shortages are constantly evolving and labor markets take time to adjust. As a result, skill shortages exhibit strong auto-regressive properties (as can be observed in Section 4.4). Therefore, we construct models to predict skill shortages which account for these temporal characteristics.

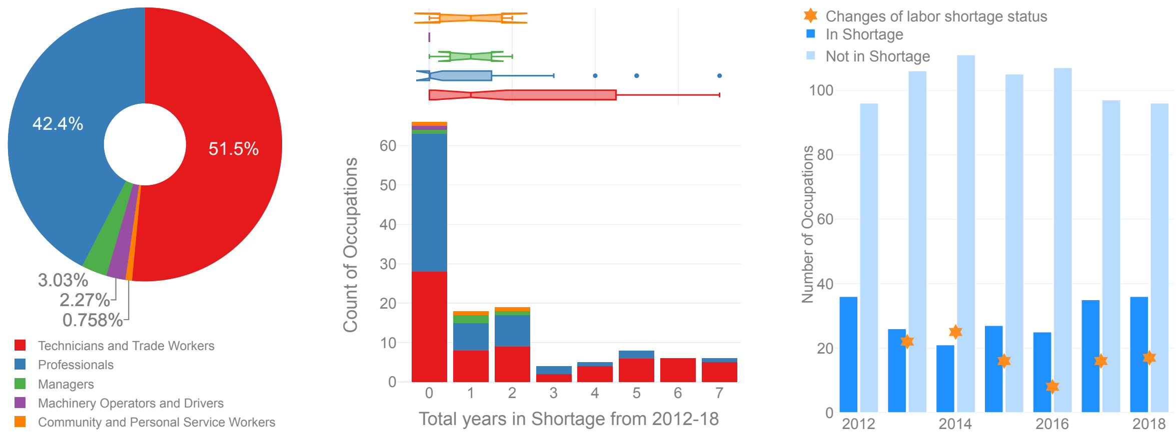
XGBoost, was not specifically built for time series prediction tasks and it makes the fundamental assumption that observations are independent. However, XGBoost has been applied for several time series prediction tasks and achieved impressive results [191, 273, 346]. We also use XGBoost to make predictions on temporal data in this research. To account for the temporal nature of skill shortages, we engineer 'auto-regressive lagged features' – i.e. for each feature included in the model, we also include its offset values over a specified number of past periods. In our experiments in Section 4.4, we use two auto-regressive lag periods. The inclusion of such auto-regressive lagged features provides each observation with temporal characteristics.

Predicting one year into the future. The dataset is organised into yearly intervals to match the ground truth. While the descriptive features are available at most three months after the year's end, the DESE skills shortage ground truth is often published 12-18 months (or longer) after the reported period. Therefore, our models are setup to predict skill shortages one year in advance of the official government release.

Training model hyper-parameters. Like most machine learning algorithms, XGBoost has a set of hyper-parameters – parameters related to the internal design of the algorithm that cannot be fit from the training data. The hyper-parameters are usually tuned through search and cross-validation. In this work, we employ a RandomizedSearch [65] which randomly selects a (small) number of hyper-parameter configurations and performs evaluation on the training set via cross-validation. We tune the hyper-parameters for each learning fold using 2500 random combinations, evaluated using a 5 cross-validation. We also implemented 'oversampling' to accommodate for the imbalance between the *In Shortage* and *Not in Shortage* classes in the ground-truth (see Section 4.4.1). This technique involves randomly duplicating observations from the minority class (*In Shortage*) and adding them to the training dataset (see Supplementary Materials (Section 4.7) for more details).

Performance measures. Here, we measure the performance of our prediction using three standard Machine Learning performance measures: precision, recall, and F1. For more details on these metrics, please refer to the Supplementary Materials (Section 4.7). In our results in Section 4.4, we report the macro-precision, macro-recall and macro-F1, which are the means of the indicators over the two classes. This makes sure that the minority class (here the *In Shortage* class) are not under-represented in the results.

Train-test split. Consistent with established Machine Learning practices, we separated the dataset into 'training' and 'testing' sets. This split was implemented temporally, with observations from 2012-2016 included in the training dataset, and observations from 2017-2018 included in the testing dataset. The training dataset consisted of 660 observations (71% of total observations) and the testing dataset consisted of 264 ob-



((a)) ((b)) ((c))

Figure 4.1: **Overview of Skills Shortage Dataset**: **(a)** Proportion of occupations represented in dataset by ANZSCO Major Group classes; **(b)** count of occupations grouped by the number of years *In Shortage* (the colors correspond to the same occupational categories observed in the Fig. 4.1(a) legend); **(c)** yearly distribution of occupations classified as *Not in Shortage* (light blue bars - 718 total) or *In Shortage* (dark blue bars - 206 total) and the yearly number of occupations whose shortage status has changed since the previous year (orange stars).

servations (29% of total observations). Segmenting the dataset into temporal training and testing sets is done to ensure objectivity in the evaluation process and reflect the temporal nature of the ground-truth.

4.4 Results

In this section, we first perform an exploratory data analysis of the constructed dataset (Section 4.4.1) before showing three sets of results that directly answer our research questions from Section 4.1. In the first set of results, we compare two methods to analyze the underlying skill demands of 'Data Scientists'. Next, we implement Machine Learning models to predict skill shortages of occupations, as outlined in Section 4.3.3. Last, we extract and analyze the feature importance data from these models to identify which variables are most predictive of skill shortages. We incorporate three data sources to construct the dataset that we use for modeling; these data sources include (1) job ads data from BGT, (2) employment statistics from ABS, and (3) occupational skills shortage classifications from the DESE, which are described in Section 4.3.1.

4.4.1 Profiling the Skills Shortage Prediction dataset

Here, we perform an exploratory data analysis and profiling of the dataset. The purpose is to understand the biases and imbalances introduced during the dataset's construction.

Construct the Skills Shortage Prediction dataset. The compiled dataset describes 132 unique occupations during the period 2012-2018. Each row consists of a tuple (occupation, year), and it describes the given occupation during that particular year using its ANZSCO identifiers, the values for each of the descriptive features (described in Section 4.3.1 and the Supplementary Materials (Section 4.7)), and the auto-regressive lagged features (described in Section 4.3.3). Our resulted dataset contains 924 occupation-year tuples (rows) described by a total of 32 features (excluding lagged feature periods). The binary target variable is its shortage status during that year: *In Shortage* or *Not in Shortage*. In constructing this dataset, we analyzed the auto-correlations within the constructed features, which are presented in the Supplementary Materials (Section 4.7). Unsurprisingly, we found that features from the same or similar categories were strongly correlated, whereas features from different datasets (job ads data and employment statistics) tended to be uncorrelated; the analysis did not yield consequential results. We next profile the contributed dataset, and we uncover a series of specifics that should be considered during the modeling process. The Skills Shortage Prediction dataset and code will be made available upon acceptance of the paper.

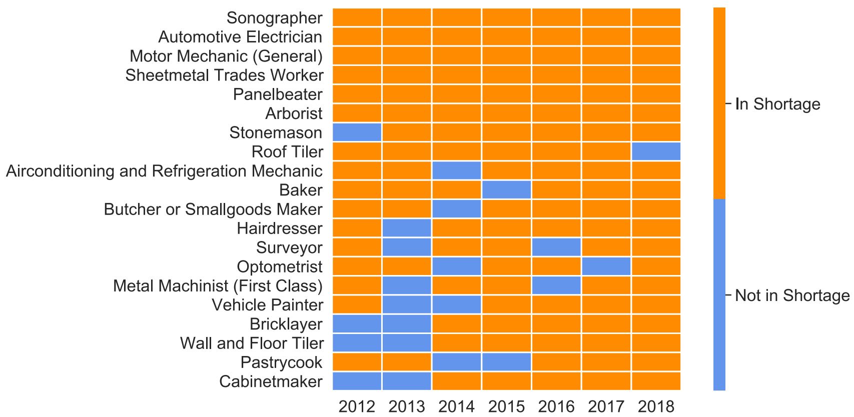
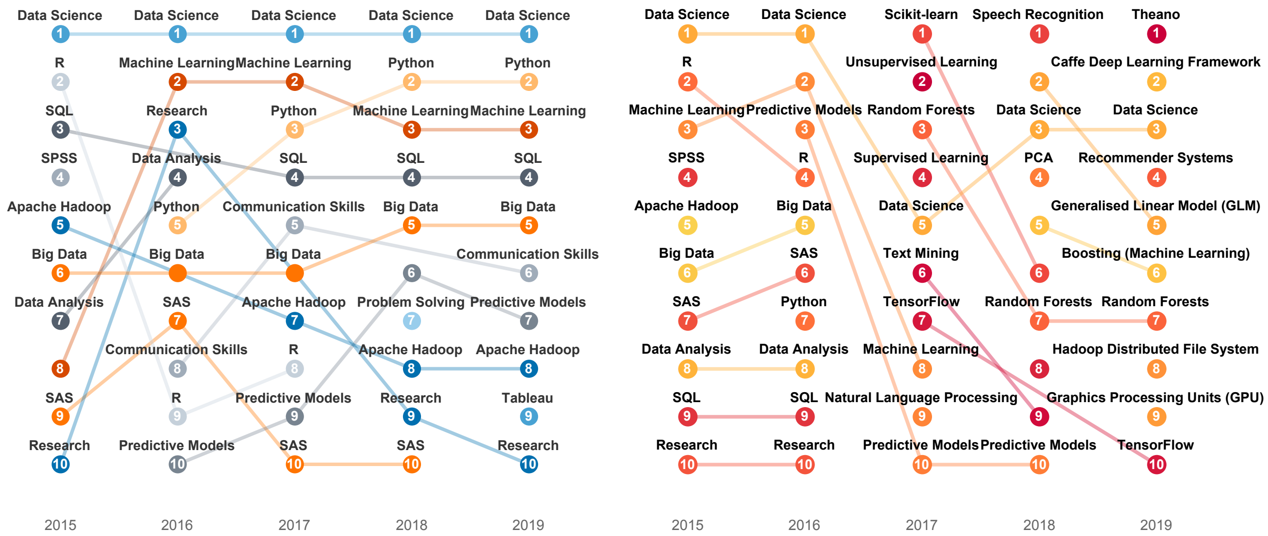


Figure 4.2: Top occupations most *In Shortage* at the ANZSCO 6-digit occupational level.

Prevalence of Technicians and Professionals. Fig. 4.1(a) shows that the occupational classes measured by the DESE disproportionately represent 'Technicians and Trades' and 'Professionals'. Collectively, these two major occupational groups account for 94% of occupations included in the dataset. This is higher than the number of workers



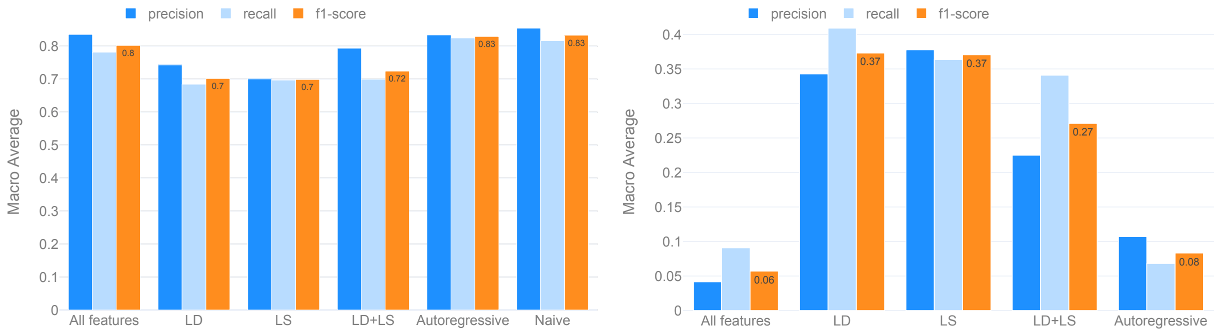
((a)) ((b))

Figure 4.3: **Comparison of two methods to analyze underlying skill demands of occupations in shortage**: **(a)** posting frequency of skills in an occupation; **(b)** Revealed Comparative Advantage of skills in an occupation to normalize highly-common skills and uncover skills most relevant to an occupation.

actually employed in these occupational classes. For instance, the ABS indicates that ‘Professionals’ represent approximately 24% of employment in Australia [31]. The bias and validity of the ground truth are discussed in Section 4.3.1.

Most occupations are *Not in Shortage*. In advanced labor markets, prolonged skill shortages are rare [85] and most occupations are ‘Not in Shortage’. This is visible in our ground truth data where there are over three times as many occupations classified as *Not in Shortage* than *In Shortage* (see Fig. 4.1(c)). However, this has important modeling implications and requires hyper-parameter tuning to sufficiently adjust for these imbalances, as discussed in Section 4.3.3.

Some occupational classes are *In Shortage* more often than others. The shortage status of occupations is updated yearly in our dataset, and occupations can be *In Shortage* for a period of time between 1 and 7 years (the extent of our dataset). In Fig. 4.1(b), we count the number of occupations *In Shortage* based on the period of time they stay in shortage, and we color them by their occupational class. We observe that the occupations belonging to the ‘Technicians and Trades’ class (shown in red) are *In Shortage* for longer periods of time than any other occupational classes, including ‘Professionals’. Furthermore, the ‘Technicians and Trades’ class makes up the majority of occupations *In Shortage* for four years or more. Generally, a small number of ‘Technicians



((a)) ((b))

Figure 4.4: **Skills Shortage prediction results**: **(a)** While the prediction results are highly auto-regressive, Labor Demand and Labor Supply features alone (and combined) perform almost as well for predicting occupational shortages; **(b)** Labor Demand and Labor Supply features perform better than other features at predicting shortage status changes of occupations.

and Trades' occupations classified *In Shortage* tend to persist over several years, further illustrated in Fig. 4.2. These finding, coupled with the fact that 'Technicians and Trades' is the largest represented class in the ground truth (see Fig. 4.1(a)) indicates that the ground-truth exhibits biases toward the 'Technicians and Trades' workers occupational class – probably due to the necessities of the Australian labor market.

Changes in labor shortage status. Changes to skill shortages of occupations are a key factor that determine adjustments to education, skilled immigration, and labor market policies. The ability to predict such yearly classification changes is therefore critical to models attempting to predict skill shortages. In Fig. 4.1(c), we count the number of occupations *In Shortage* and *Not In Shortage* per each calendar year between 2012 and 2018, alongside with the number of occupations that flip their shortage status (from *In Shortage* to *Not In Shortage*, or the other way around, shown by the orange hexagrams). We see that changes to occupational skills shortage status are relatively rare (about 20 or less occupations change their status every year). This suggests that the ground-truth contains auto-regressive properties – i.e., the status this year is most likely the same as last year – which is an important modeling consideration, particularly for predicting shortage changes.

4.4.2 Skill importance levels for Data Scientists.

Here, we compare the two approaches to assess skill importance for occupations that we introduced in Section 4.3.2, and we showcase them for the occupation 'Data Scientist'.

The first approach is to perform temporal skill counts grouped by occupation (or another grouping source). Fig. 4.3(a) highlights the top 10 skills for 'Data Scientist' obtained using this approach, for each between 2015 and 2019. Visibly, skills like 'Communication Skills', 'Research', and 'Problem Solving' regularly rank in the top 10, however, these are among the most common skills in the BGT dataset – for example, 'Communication Skills' is present in over one-quarter of all job ads. This is because skill counts do not normalize for highly common skills that are present in all or most occupations, making it questionable whether this can be used as a proxy for skill importance within an occupation.

The second approach detailed in Section 4.3.2 is the RCA approach. Fig. 4.3(b) shows the top 10 yearly skills obtained using RCA. Visibly, the obtained top skills are considerably more specific to the 'Data Scientist' occupation. Machine Learning and Deep Learning tools and techniques dominate the ranked list, while some core Data Science skills seen in Fig. 4.3(a) remain. This method also captures the rise of emerging skills (such as *Generalized Linear Models*, *Boosting* or *Random Forests*), which are critical for occupations in shortage.

4.4.3 Predict Skill Shortages

Here, we detail the results of two predictive exercises. First, we predict the shortage status of occupations and we perform an ablation study to identify the most important sets of features. Second, we show the results of the more difficult task of predicting shortage status changes (when an occupation flips its shortage status between *In Shortage* and *Not In Shortage*).

Predict occupation shortages. We predict occupation shortages following the setup described in Section 4.3.3. We equally study which class of features is most predictive by performing an ablation study – i.e. we repeat the predictive experiment multiple times using all the features, or only subsets of features. We train and evaluate the following feature input configurations: **All-In**: all features included; **LD**: labor demand features included only; **LS**: labor supply features included only; **LD + LS**: labor demand and labor supply features included; **Auto-regressive Predictor**: lagged target features included only; **Naive Predictor**: copy target variable from the previous time period. Fig. 4.4(a) shows the prediction performance – macro- precision, recall and F1 (higher is better) – of the different setups. Due to the strong auto-regressive properties of the problem, the *Naive Predictor* and the *Auto-regressive Predictor* achieve the highest performance ($F1 = 83\%$), however they always predict the last shortage status for each occupation. These predictors are useless for occupations that flip their status, which are of strong interest in real-world applications. Visibly, the models that exclude the auto-regressive features (i.e. the LD and/or LS models), maintain solid performance levels (up to F1=72%). The significance of this finding is discussed in Section 4.5.

Predict shortage status changes. We evaluate the same classifiers trained as described above on a slightly different problem: how well can they predict the *changes in shortage status*? To achieve this, we filtered occupations in the testing dataset to include only those with a different skills shortage classification to the previous year. For example, as 'Architects' were classified as *In Shortage* in 2017 but were *Not in Shortage* in 2016, they were therefore included in the performance evaluation. Fig. 4.4(b) shows the resulting prediction performance: precision, recall and F1. Visibly, the performances decreased substantially, and the hardest hit are the models leveraging the auto-regressive property (with *Naive* obtaining zero everywhere). The reason is that shortage status changes are fairly rare events, (see Fig. 4.1(c)) which auto-regressive classifiers completely miss. The highest performing models use LD or LS features. This is particularly relevant to real-world scenarios, where researchers closely follow occupations that change their status, as this has policy and immigration implications.

4.4.4 Feature Importance for Predicting Skill Shortages

As seen in Fig. 4.4(a), the model with the auto-regressive features has the highest performance, so previous shortage classifications are the most important features for predicting skill shortages in this dataset. However, longitudinal datasets of skill shortages, like the data used for this analysis, are rare. Therefore, auto-regressive target features are often unavailable for analyzing skill shortages in other labor markets. Labor demand and labor supply features, however, are standard and available across most labor markets. Here, we conduct feature importance analysis on the 'LD + LS' model seen in Fig. 4.4(a) in order to draw insights into which of these features are most predictive of skill shortages. We use the 'Gain' metric, which shows the relative contribution of each feature to the model by calculating the features' contribution for each tree in the XGBoost model. A higher gain score indicates that a feature is more important for generating a prediction.

Fig. 4.5 shows that variations of the labor supply feature 'Hours Worked' are the most predictive for skill shortages, as they account for 6 of the top 20 most important features – see positions 1, 2, 4, 10, 12, 15 in Fig. 4.5. The next most important features belong to the labor demand class. Namely, years of 'Education' and 'Experience' demanded by

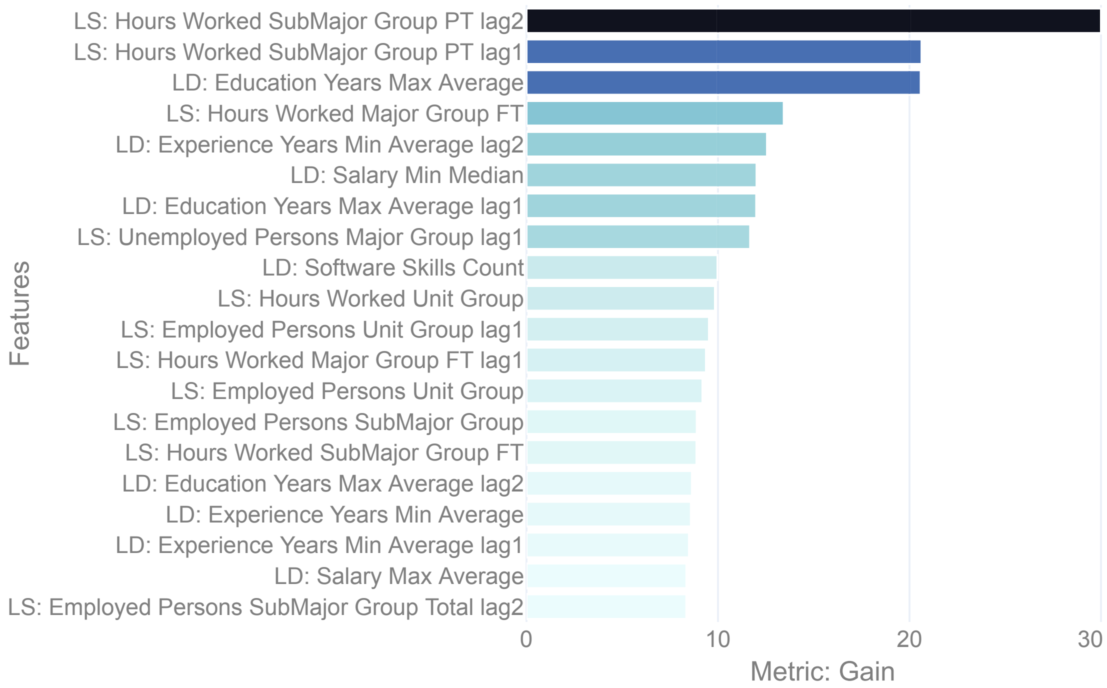


Figure 4.5: Feature importance of Labor Demand (LD) and Labor Supply (LS) feature model.

employers and median 'Salary' levels in job ads. A brief interpretation of these feature importance levels follows in Section 4.5.

4.5 Discussion

Trade-off between performance and data availability. The highest performing models in Fig. 4.4(a) exhibit strong auto-regressive properties. This is to be expected given that changes in the skill shortage status of occupations tend to be rare, as seen in Figs. 4.1(c) and 4.2. However, removing the auto-regressive target features and leaving the labor demand and labor supply features maintains a relatively strong result (F1=72%). This is significant because labor demand and labor supply data sources are available across multiple labor markets, whereas longitudinal skill shortages data at the occupational level are rare in most labor markets. This suggests that while labor demand and labor supply data contain rich information for detecting skill shortages, there is a trade-off between prediction performance and data availability when deploying in new labor markets.

Auto-regressive features cannot predict shortage status changes. Given the strong auto-regressive nature of skill shortages (i.e. the best indicator of an occupation being in shortage this year is if it was in shortage last year), classifiers have the tendency of over-leveraging the information from the past. While this may help performance indicators, it considerably reduces the value of the prediction in a real-life setups. Shortage status changes (when an occupation moves between *Not In Shortage* and *In Shortage*) have policy and immigration implications, as governments decide skilled immigration rules based on the needs of the labor market. In other words, it is more important to be able to predict when an occupation shortage *status changes* than simply predicting its next status. Visibly in Fig. 4.4(b), the performances of the classifiers leveraging autoregressive features are significantly reduced when predicting shortage status changes. Nonetheless, we found that labor demand and labor supply data were most predictive of shortage changes, respectively. This is significant because it further highlights the value of near-real time data sources (job ads data) and freely available data sources (employment statistics). Both labor demand and labor supply data sources could be leveraged to replicate our modeling approach in other markets to assist policy-makers to better preempt skill shortage changes of occupations. This could help with critical tasks such as forward planning for education and training policies, skilled immigration, and workforce transitions.

Understanding what matters for predicting skill shortages. The most important features from the 'LD + LS' model (seen in Fig. 4.5) are consistent with the literature on skill shortages [85, 105, 120, 180, 262, 287]. Specifically, 'Hours Worked' is considered an important indicator for occupations in shortage [180, 287] due to the following rationale: when a shortage exists for an occupation, the demands placed upon workers classified in that occupation are naturally high, which manifests in higher work intensity and longer work hours. This is reflected in Fig. 4.5 where the 'Hours Worked' variables are represented in 6 of the top 20 most important features.

With regards to labor demand, years of 'Education', years of 'Experience', and median 'Salary' are all highly important features for predicting occupational skill shortages. This is consistent with prior work [120], which shows that when an occupation is in shortage, employers adjust job requirements to try and fulfill their demands. With regards to these features, this typically involves lowering the requirements of education and experience and increasing salary levels to attract more candidates.

4.6 Conclusion and Future Work

In this research, we (1) compared two methods to analyze the skill demands of occupations in shortage; (2) we constructed a Machine Learning framework to predict temporal skill shortages of occupations; and (3) we analyzed feature importance data to understand which labor supply and labor demand features are most predictive of occupational skill shortages. The methods and findings from this work can assist policy-makers to better measure and predict skill shortages of occupations. Similarly, educators could apply this work to better identify market demands and adjust their curricula accordingly.

The biggest limitation with skills shortage research is the lack of representative data at the occupational level. The 'Historical List of Skills Shortages in Australia', used in this research, is among the world leaders in this regard, despite its shortcomings discussed in Section 4.3. Therefore, systematically measuring occupational skill shortages is arguably the most important work that can be done to advance the knowledge of skill shortages. Other future work could apply the framework we have developed predict skill shortages in other labor markets. Additionally, different features could be constructed as descriptive variables, and more auto-regressive lag periods could be considered. Lastly, another research avenue could assess how these results could be improved by applying other predictive tools, such as Deep Learning approaches.

4.7 Supplementary Materials

This document is accompanying the submission *Predicting Skill Shortages in Labor Markets: A Machine Learning Approach*. The information in this document complements the submission, and it is presented here for completeness reasons. It is not required for understanding the main paper, nor for reproducing the results.

4.7.1 Cyclical and Structural Factors Affecting Skill Shortages

Macroeconomic cycles can affect skill shortages. During periods of economic expansion, skill shortages tend to increase as firms seek to hire skilled labor to meet new and growing market demands [85]. The 'Manpower Talent Shortage Survey' [224] is the largest skill shortage survey in the world. The global survey found that skill shortages have increased from 30% in 2009 to 45% in 2018, equating to a 12 year high. Similarly, the annual Cedefop skills mismatch survey in Europe [105] found that labor market shifts in the aftermath of the economic crisis have resulted in the stated inability of employers to fill their vacancies with suitably skilled workers.

Structural changes to labor markets also influence skill shortages. These most notably take the form of demographic changes, technological advances, and globalization. Demographic changes affect the demand for goods and services. For instance, as the average age of a population increases, so does their demand for healthcare services. This subsequently increases the aggregate labor demand for workers with healthcare related skills [264]. As the average age is increasing for almost all advanced economies [343], these structural demographic changes are likely to affect skill shortages for specific occupational classes, such as healthcare services.

Technological advances introduce structural changes that can exacerbate skill shortages. As firms adopt new technologies, they seek skilled labor to implement and make productive use of these new technologies. This can create dynamics of 'skill biased technological change' [7, 199], whereby the acceleration of demand for technical skills outweighs the available supply of workers who possess such skills. There is evidence of these dynamics currently occurring as a result of the growing demands for Data Science and Machine Learning skills [120]. While the capacity to collect, store, and process information may have sharply risen, it is argued that these advances have far outstripped present capacities to analyze and make productive use of such information [184]. Claims of Data Science and Advanced Analytics (DSA) skill shortages are being made in labor markets around the world [70, 213, 225]. Two studies conducted using job ads data assessed DSA labor demands and the extent of skill shortages. The first was an industry research collaboration between Burning Glass Technologies (BGT), IBM, and the Business-Higher Education Forum in the US [227]. The research found that in 2017 DSA jobs earned a wage premium of more than US$8,700 and DSA job postings were projected to grow 15% by 2020, which is significantly higher than average. In another study commissioned by the The Royal Society UK [70], job ads data were analysed for DSA jobs in the UK. The results again also showed high and growing levels of demand for DSA skills (measured through posting frequency) and wage premiums for DSA related occupations.

Globalization can act as a shock to labor markets that induce or deepen skill shortages. The offshoring of labor tasks can increase the polarization of labor markets by reducing the domestic demand for middle-skilled jobs [85]. This causes a process of labor reallocation, as workers attempt to transition between jobs. If the reallocation of labor is inefficient, skill shortages can increase because the supply of skilled workers is insufficient to meet the evolving labor demands of growing sectors.

The current work proposes a robust data-driven method that assesses skill shortages and that uses machine learning to account for the factors that affect skill shortages.

4.7.2 Oversampling

Oversampling is a technique that involves randomly duplicating observations from the minority class (*In Shortage* in the case of this research) and adding them to the training dataset. The main benefit of oversampling is that it creates a balanced distribution of target variables without 'data leakage' that occurs from 'under-sampling' (that is, randomly removing observations from the majority class). Creating a balanced distribution of predictive classes is particularly important for a range of classification algorithms [78]. However, a shortcoming of oversampling is that it can increase the likelihood of overfitting, as exact copies of the minority class are constructed [143]. The oversampling ratio is defined as:

$$Oversampling\ Ratio = \frac{\sum(Majority\ Class)}{\sum(Minority\ Class)}$$

The output of this ratio was specified as a hyper-parameter value in each model type that we constructed.

4.7.3 Performance metrics

Precision measures how many of the predictions were correct. Recall measures the completeness of the prediction – how many of the true answers were correctly uncovered. The F1 is the harmonic mean of precision and recall – a classifier needs to achieve both a high precision and a high recall in order to obtain a high F1. Formally, these are defined as:

$$Precision = \frac{TP}{TP+FP};$$

$$Recall = \frac{TP}{TP+FN};$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

where TP are the number of true positives – number of correctly identified items of the class of interest; FP are false positives (items incorrectly predicted as pertaining to the class of interest); and FN are false negatives (items incorrectly predicted as not being of interest). Note that one can compute the precision, recall and F1 for each class of interest (here, both the *In Shortage* and the *Not in Shortage*), and the scores for each class could be wildly different as one class might be more predictive than the other.

4.7.4 Using a standardized occupation taxonomy – ANZSCO.

All data sources mentioned above correspond to their respective occupational classes according to the Australian and New Zealand Standard Classification of Occupations (ANZSCO). [25] ANZSCO provides a basis for the standardized collection, analysis and dissemination of occupational data for Australia and New Zealand. The structure of ANZSCO has five hierarchical levels - major group, sub-major group, minor group, unit group and occupation. The categories at the most detailed level of the classification are termed 'occupations'. Depending on data availability, labor statistics were included in the models from the occupation level through to the major group level.

There are some significant shortcomings to analyzing occupations within ANZSCO classifications. Official occupational classifications, like ANZSCO, are often static taxonomies and are rarely updated. They therefore fail to capture and adapt to emerging skills, which can misrepresent the true labor dynamics of particular jobs. For example, a 'Data Scientist' is a relatively new occupation that has not yet received its own ANZSCO classification. Instead, it is classified as an 'ICT Business & Systems Analyst' by ANZSCO, grouped with other job titles like ‘Data Analysts’, ‘Data Engineers’, and ‘IT Business Analysts’. However, as ANZSCO is the official and prevailing occupational classification system, all data used for this research are in accordance with the ANZSCO standards.

4.7.5 Summary of constructed features

Table 4.1: Summary of constructed features and their explanation.

	Name	Meaning and explanation
Labour Demand	Posting Frequency:	number of job advertisement vacancies
	Max Median Salary:	maximum median salary advertised
	Min Median Salary:	minimum median salary advertised
	Max Average Salary:	maximum average salary advertised
	Min Average Salary:	minimum average salary advertised
	Max Average Experience:	maximum average years of experience required
	Min Average Experience:	minimum average years of experience required
	Max Average Education:	maximum average years of formal education required
	Min Average Education:	minimum average years of formal education required
	Specialised Count:	total count of required skills considered specialised to a specific vocation
	Baseline Count:	total count of skills that are considered applicable across vocations
	Software Count:	total count of skills that are software-related
Labour Supply	Unit Total Employed:	total number employed at ANZSCO Unit level (000’s)
	Unit Total Hours Worked:	total hours worked at ANZSCO Unit level (000’s)
	Sub FT Employed:	total employed full-time at ANZSCO Sub-Major level (000’s)
	Sub PT Employed:	total employed part-time at ANZSCO Sub-Major level (000’s)
	Sub Total Employed:	total employed at ANZSCO Sub-Major level (000’s)
	Sub FT Hours Worked:	total full-time hours worked at ANZSCO Sub-Major level (000’s)
	Sub PT Hours Worked:	total part-time hours worked at ANZSCO Sub-Major level (000’s)
	Sub Total Hours Worked:	total hours worked at ANZSCO Sub-Major level (000’s)
	Major FT Employed:	total employed full-time at ANZSCO Major level (000’s)
	Major PT Employed:	total employed part-time at ANZSCO Major level (000’s)
	Major Total Employed:	total employed at ANZSCO Major level (000’s)
	Major FT Hours Worked:	total full-time hours worked at ANZSCO Major level (000’s)
	Major PT Hours Worked:	total part-time hours worked at ANZSCO Major level (000’s)
	Major Total Hours Worked:	total hours worked at ANZSCO Major level (000’s)
	Major Unemployed FT Seekers:	total unemployed seekers full-time at ANZSCO Major level (000’s)
	Major Unemployed PT Seekers:	total unemployed seekers part-time at ANZSCO Major level (000’s)
	Major Unemployed Total Seekers:	total unemployed seekers at ANZSCO Major level (000’s)
	Major Total Weeks Searching:	total number of weeks unemployed persons job searching at ANZSCO Major level (000’s)
	Major Underemployed Total:	total number of persons underemployed at ANZSCO Major level (000’s)
	Major Underemployed Ratio:	ratio of underemployed persons at ANZSCO Major level

4.7.6 Feature correlation analysis

During the exploratory data analysis stage of this research, we conducted a feature correlation analysis. Unsurprisingly, features from the same or similar categories were strongly correlated, whereas features from different datasets tended to be uncorrelated. Therefore, this analysis did not yield consequential results.

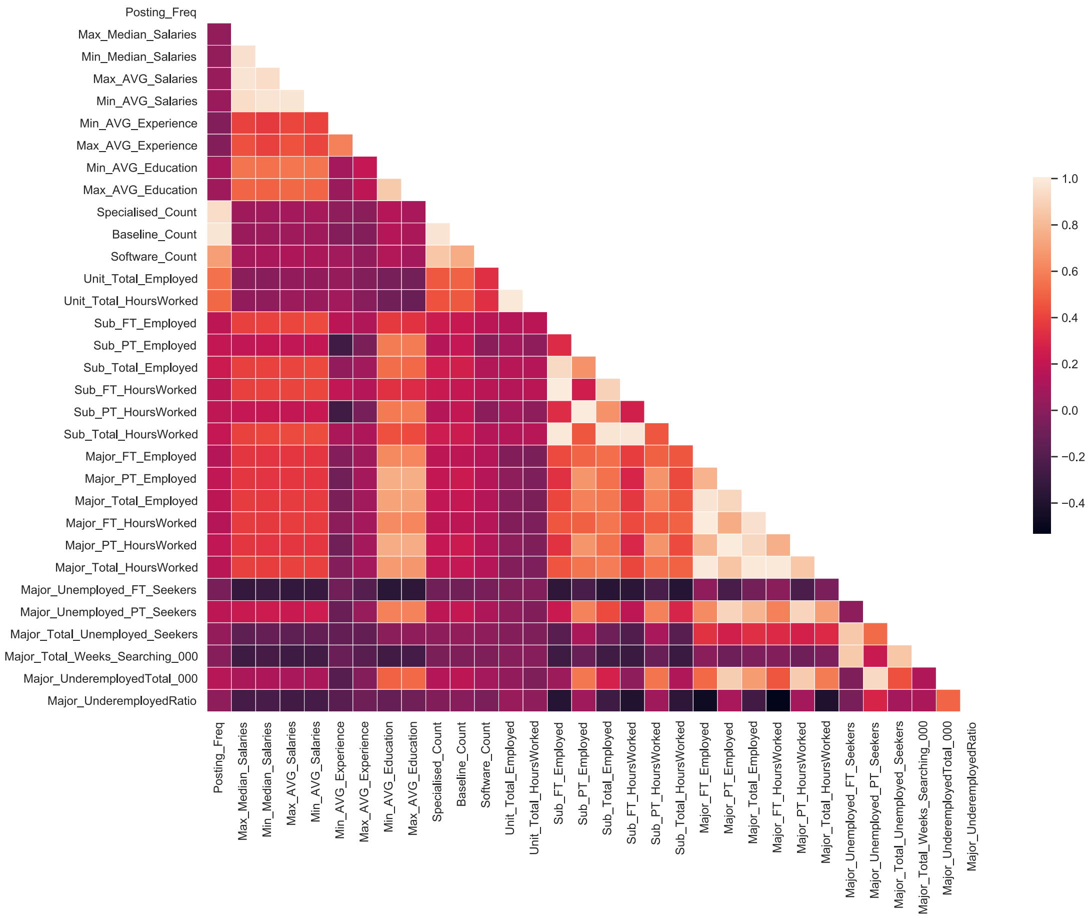


Figure 4.6: Correlation analysis between modeled features.

Chapter 5

Paper 3 - Layoffs, Inequity and COVID-19: A Longitudinal Study of the Journalism Jobs Crisis in Australia from 2012 to 2020

Preamble

This paper provides an in-depth case study of journalism jobs in Australia from 2012-2020. The journalism industry has experienced significant disruption following the advent of digital platforms and new technologies. As a result, many journalists have been forced to reconsider their career paths and transition to new occupations to find work. This research combines a range of labour market data sources and applies data science techniques to provide a comprehensive assessment of journalism jobs in Australia from 2012-2020. The results from this paper contribute to the broader theme of job transitions for this thesis. The paper provides tools and techniques to analyse job transitions from labour market data; it also contributes among the most comprehensive accounts of the journalism job market in Australia since 2012. This research paper was published in the *Journalism* journal in February 2021 [118].

Abstract

In Australia and beyond, journalism is reportedly an industry in crisis, a crisis exacerbated by COVID-19. However, the evidence revealing the crisis is often anecdotal or limited in scope. In this unprecedented longitudinal research, we draw on data from the Australian journalism jobs market from January 2012 until March 2020. Using Data Science and Machine Learning techniques, we analyse two distinct data sets: job advertisements (ads) data comprising 3,698 journalist job ads from a corpus of over 8 million Australian job ads; and official employment data from the Australian Bureau of Statistics. Having matched and analysed both sources, we address both the demand for and supply of journalists in Australia over this critical period. The data show that the crisis is real, but there are also surprises. Counter-intuitively, the number of journalism job ads in Australia rose from 2012 until 2016, before falling into decline. Less surprisingly, for the entire period studied the figures reveal extreme volatility, characterised by large and erratic fluctuations. The data also clearly show that COVID-19 has significantly worsened the crisis. We then tease out more granular findings, including: that there are now more women than men journalists in Australia, but that gender inequity is worsening, with women journalists getting younger and worse-paid just as men journalists are, on average, getting older and better-paid; that, despite the crisis besetting the industry, the demand for journalism skills has increased; and that, perhaps concerningly, the skills sought by journalism job ads increasingly include social media and generalist communications.

5.1 Introduction

Globally, the news about the news is not good. This was true before 2020, but COVID-19 has only made matters worse. Take Australia. In March 2020, newswire service the Australian Associated Press announced it would be shutting down its operations after 85 years [294]. In June, a last-minute consortium of investors and philanthropists saved the day – but salvation was merely partial, with only 85 of the company's 180 journalists, photographers and other staff retained [332]. Meanwhile, News Corp has been closing scores of regional titles (see below). In the US, the news about the news is just as bad, if not worse. In February, the country's No. 2 newspaper chain (McClatchy) declared bankruptcy [64]. Amid widespread pay cuts, furloughs and layoffs, US newsrooms reportedly shed more than 11,000 jobs in the first half of 2020 [338].

Even before COVID-19, digital technology upended journalism's advertising-driven business model [6]. As the Nieman Lab notes:

> The Internet has brought forth an unprecedented flowering of news and information. But it has also destabilised the old business models that have supported quality journalism for decades. Good journalists across the country are losing their jobs or adjusting to a radically new news environment online [209].

But is journalism *in crisis*? A wealth of research in Australia, the US and comparable countries suggests yes. Profits have been hard, if not impossible, to come by; many firms were struggling or collapsing; and layoffs and redundancies were the norm [6]. As Fenton [142] wrote in a paper centred on the UK, 'News media are in crisis. The crisis is being managed by closing papers or shedding staff [and] these cuts are having a devastating effect on the quality of the news.' That was nearly a decade ago. Subsequent research suggests the situation has worsened significantly. In Australia, the commonly cited figure based on research by the journalists' union is that 3,000 journalism positions have been lost since 2011 [288]. For instance, it is estimated that in 2011 news publisher Fairfax Media employed about 1,000 editorial staff across the *Sydney Morning Herald*, *The Age*, *The Australian Financial Review*, and its Sunday papers, *The Sun Herald* and *The Sunday Age*. By mid-2017, however, half of those jobs were gone [348] (including the job of one of this paper's authors). And then the coronavirus wielded its scythe. Already, the reported impact of COVID-19 on journalism jobs has been devastating, with widespread closures and job losses, particularly in regional areas [113].

This research aims to assess the extent of the claimed 'journalism crisis' in Australia by analysing labour market data from January 2012 to March 2020. To fulfil this research aim, we performed a quantitative analysis of two longitudinal data sets: job advertisements (ads) for journalism jobs and the official Australian employment statistics. This allowed us to measure longitudinally the demand for and supply of journalism jobs in Australia. Further, the breadth and detail of these data provided us with the opportunity to comprehensively assess the quality and characteristics of these journalism jobs. Not only did we examine how key features of journalism jobs have changed – such as salaries, location or years of experience – but also how journalism skills required in Australia have evolved. Additionally, the available data enabled us to measure the early effects of COVID-19 on journalism jobs.

Our findings confirm that there is a crisis in Australian journalism; a crisis that appears to have worsened during the early stages of the COVID-19 pandemic. However, the data also yields more granular findings, including three surprise findings. The first finding is that advertised journalism jobs only started to decline from 2016, not before. The second finding is that as the journalism jobs market became more volatile, gender inequity worsened: women journalists who remained were younger and worse paid than the men. And the third finding is that, according to our skill similarity calculations, generalist skills such as ‘Communications’, ‘Public Relations’, and ‘Social Media’ became more important to journalism, as opposed to traditionally specialist journalism skills such as ‘Reporting’, ‘Editing’, and ‘Investigative Journalism’. These findings, together with others, reveal that the crisis in journalism is not only real, but in some ways more complex than was previously understood.

By implementing a data-driven methodology, we provide a comprehensive and longitudinal assessment of journalism jobs in Australia from January 2012 to March 2020. We tease out granular and specific trends, including the early impacts of the COVID-19 pandemic, the contrasting effects on regional and urban journalism jobs, and the gendered nature of ongoing impacts. And we analyse the underlying skills data to identify the skills sought in journalism jobs, and where people with journalism skills are likely finding alternate career paths.

5.2 Related Work & Background

Journalism jobs in crisis. If there is a crisis, the simple explanation is the Internet. (Putting aside COVID-19, to which we will return.) While digital channels have given journalism bigger audiences, they have also strangled income. Once, advertising funded journalism, but now advertising has largely migrated online. As the Australian Competition and Consumer Commission (ACCC) found in 2019, in the Final Report of its Digital Platforms Inquiry, ‘The reduction in advertising revenue over the past 20 years, for reasons including the rise of online advertising, appears to have reduced the ability of some media businesses to fund Australian news and journalism’. The ACCC cited Census data showing that ‘from 2006 to 2016, the number of Australians in journalism-related occupations fell by 9 per cent overall, and by 26 per cent for traditional print journalists (including those journalists working for print/online news media businesses)’. Further, the ACCC cited data provided by leading media companies showing that the number of journalists in traditional print media businesses fell by 20 per cent from 2014 to 2018 ‚Äì a time of growth for Australia's population and economy [6].

However, the pressures on news media were not spread evenly. For instance, local news in particular bore the brunt. Between 2008 and 2018, 106 local and regional newspaper titles closed across Australia, representing a 15 per cent decrease in the number of such publications. As a result, 21 local government areas previously served by a newspaper were now without coverage, including 16 local government areas in regional Australia [6]. These figures are mirrored in the US. In 2018, Abernathy [5] from the Hussman School of Journalism and Media at UNC released a report, '*The Expanding News Desert*', which found that the US had lost almost 1800 papers since 2004, with 7112 remaining (1283 dailies and 5829 weeklies). This meant that the US lost roughly 20 per cent of its newspapers between 2004 and 2018. These closures included large dailies such as the *Tampa Tribune* and the *Rocky Mountain News*, but also many newspapers that had circulations of fewer than 5000 and served small, impoverished communities.

As the above research reveals, news media companies were under pressure, and journalism jobs were being cut. There was some hope in the shape of new players entering the market and hiring journalists, including digital natives such as Vice and Buzzfeed. However, in 2019 these two companies were among the many that announced significant staff layoffs [162]. Worse, in 2020 Vice cut a further 155 jobs and Buzzfeed furloughed many of its workers without pay [190]. Furthermore, as Australia's ACCC notes, these publications 'tend to employ relatively few journalists' [6]. Even accounting for new arrivals, the number of journalism jobs in Australia continued to fall (our own analysis in Section 5.4 also shows this trend), and as a result there were areas (including local government, local court, health and science issues) that journalism no longer covered adequately [6].

Further research has also revealed a clearer profile of the typical journalist, and also the typical journalist who loses his/her job. Drawing on 2017 data, one study found that journalism jobs internationally were largely filled by a young, inexperienced and itinerant workforce [193]. Meanwhile, research suggests that it was journalists with extensive experience losing their jobs (at least in Australia) [304]. And those who lost their jobs faced decidedly uncertain futures. In longitudinal research tracking the post-journalism careers of Australian journalists who had been made redundant, many of those surveyed revealed they were experiencing job precarity [348]. Further, a significant minority had moved into strategic communications or public relations [348]. However, the flow of journalists into PR (and sometimes back again) is not new [97, 144, 216, 217] and our analysis also supports these previous results.

The nature of 'journalism work' has also changed. Increasingly, scholars have sought to theorise journalism in terms of boundaries and blurring [99, 214, 215, 268]. The idea of blurred boundaries is intended to capture the ways in which journalism is increasingly difficult to define, and how traditional notions of journalism have been upended in the digital age [214]. Empirical work suggests that journalism is a fluid concept that now means many things, and that the definition of journalism is changing over time [72]. For example, many contributors to social networks, including Instagrammers, can be considered to be creating work that is journalism [215]. As such, there is now no such thing as a typical journalist; rather, journalism is marked by diversity and heterogeneity rather than any unifying concept [72]. The notion of blurred boundaries aligns with our findings regarding the way journalism jobs, and journalism skills, have been shifting. Indeed, Carlson [98] argues that journalism is uncertain, which means that scholars and audiences need to work towards clarifying both the value of journalism, and its meaning. Our research has been data-driven, analysing journalism jobs data according to the Australian occupational classifications of journalists (see Supplemental Materials) and based on their underlying skills in job ads data. Nonetheless, we suggest that our findings, coupled with previous research, have the potential to further inform how precisely journalism jobs in Australia have changed during this tumultuous period for the media industry.

The impacts of COVID-19. There is a growing body of research into the impacts of COVID-19 on news and its audiences. Unsurprisingly, the research reveals that the outbreak of the global pandemic was accompanied by a marked upswing in news consumption in Australia [272], the US [102] and the UK [197]. In Australia in 2019, 56 per cent of Australians accessed news more than once a day; by April 2020, three months after the first local case of COVID-19 was confirmed, that figure had jumped to 70 per cent [272]. Among other things, this increase involved audiences returning to television and legacy media in greater numbers [197, 272]. Soon, however, many people started avoiding news, and especially news about coronavirus - because it made them anxious [197, 272]. As Kalogeropoulos et al. [197] wrote following a survey of UK audiences conducted in May, 'After an initial surge in news use, there has been a significant increase in news avoidance.'

Ultimately, COVID-19 gave rise to a paradox. The above surveys show that, as audiences sought out information to stay safe, there was a dramatic surge in the consumption of news - at least initially. At the same time, however, news outlets found it even harder to make money, as advertising dried up even further [133, 267, 285]. With concerts cancelled and restaurants shuttered, promoters and restaurateurs had nothing to advertise, and the impacts on local and regional news were especially harsh [133]. On March 25, 2020, *The Atlantic* ran a story under the headline, 'The coronavirus is killing local news' [333]. The author urged people to subscribe: 'Among the important steps you should take during this crisis: Wash your hands. Don't touch your face. And buy a subscription to your local newspaper.' As one US media expert noted in late March, 'Advertising, which has been doing a slow disappearing act since 2008, has been cut in half in the space of two weeks' [133].

Even before COVID-19, the advertising crisis for journalism has been described not as a single black swan, but as a flock of black swans [133]. By one estimate, from 2006 to 2020, US newspapers lost more than 70 percent of their advertising dollars [133]. COVID-19 further cruelled advertising, compounding the strain on news media and the journalists they employ [267].

In Australia, there were widespread closures and job losses before the pandemic, but COVID-19 compounded the problem. In late March, Rupert Murdoch's publishing business News Corp warned of 'inevitable' job cuts and the closure of regional titles [233]. Soon afterwards, News Corp - Australia's biggest publisher - suspended the print editions of 60 Australian newspapers, including the *Manly Daily* and *Wentworth Courier* in Sydney, the *Brisbane News* and the *Mornington Peninsula Leader* in Victoria [233]. In May, News Corp confirmed that more than 100 of its local and regional mastheads would either switch to digital only or disappear completely [232]. These cuts came in the wake of a dramatic drop in advertising from the entertainment, restaurant and real estate industries, the titles' main revenue sources. The global pandemic is ongoing, and its lasting impact on journalism remains to be seen. Our findings, drawn from data that runs until March 2020, are early and indicative rather than definitive.

Job ads as a proxy for labour demand. Job ads provide 'leading' indicators of shifting labour demands as they occur, as opposed to the 'lagging' indicators from labour market surveys. Consequently, job ads are increasingly used as a data source for analysing labour market dynamics [70, 227]. For instance, job ads data have also been used to assess labour shortages. Dawson et al. [120] defined a range of indicators to evaluate the presence and extent of shortages, such as posting frequency, salary levels, educational requirements, and experience demands. They also built a metric based on the forecasting error from Machine Learning models trained to predict posting frequency. Intuitively, occupations experiencing high posting volatility are difficult to predict. Subsequent work showed these indicators to be predictive of labour shortages in the Australian Labour Market [121]. In the present research, in Section 5.4, we use a similar set of indicators to analyse labour demand for journalists. Further details on job ads data are provided in the Supplemental Material.

Analysing journalism jobs. Journalism jobs have previously been analysed using job ads. Young and Carson [345] collected and assessed how Australian media outlets defined journalism job positions when hiring journalists from November 2009 to November 2010. The authors used a content analysis methodology and manually labelled data fields, such as employer, educational qualifications, job responsibilities, experience requirements, location, work hours, media platform, skill demands, job title, and any other miscellaneous information. The authors found that journalism was not a high priority during this period; instead employers advertised four times as many job ads for sales, marketing, and advertising positions.

More recently, Guo and Volz [170] conducted content analysis on 669 journalist job announcements from US media organisations from 1 July to 31 December 2017. The authors' objective was to define, compare, and analyse the journalists' expertise requirements as expressed through job ads. To achieve this objective, the authors manually reviewed and codified job vacancies. This research found that 'multi-skilled' journalists are experiencing higher levels of demand. The authors also found that journalists' ability to flexibly adapt to changing situations was a characteristic of growing importance. These studies, while significant, are relatively limited in scope. In this paper, we analyse a nine-year dataset of job ads which allows us to uncover longitudinal dynamics of journalism jobs.

Historic employment levels of journalists in Australia have also been analysed by O'Regan and Young [268]. The authors used five-yearly census data and found that not only has the advent of digital platforms coincided with the decline of many types of journalists (for example, 'Print', 'Radio', 'Television' and 'Editors'), but employment has shifted into related professions, such as 'Authors' and 'Public Relations'. O'Regan and Young's paper built on earlier research by Higgs and Cunningham [186]. Our research complements the findings of O'Regan and Young [268], providing additional labour demand detail from job ads data while also matching it with labour supply data from employment statistics.

Limitations of job ads data. Job ads data are an incomplete representation of labour demand. Some employers use traditional forms of advertising for vacancies, such as newspaper classifieds, their own hiring platforms, or recruitment agency procurement. Furthermore, anecdotal evidence reveals that some vacancies are filled informally, using channels such as word of mouth, professional networks and social media. Job ads data also over-represent occupations with higher-skill requirements and higher wages, colloquially referred to as 'white collar' jobs [101]. Finally, just because a job is advertised, does not mean that the position will be, or has been, filled. Despite these shortcomings, job ads provide extremely rich information for what employers are demanding in near real-time; including information that cannot be gathered from employment statistics. Given the sample size of journalism job ads available and the detailed skills extracted in the data set, we are confident that the journalism job ads used for this research provide a useful indication of journalism labour demand.

Employment statistics and occupational standards. Employment statistics provide data on populations employed in standardised occupational classes. Occupations in Australia correspond to their respective occupational classes according to the Australian and New Zealand Standard Classification of Occupations (ANZSCO) [25].

There are significant shortcomings to analysing occupations within ANZSCO categories. Official occupational taxonomies (like ANZSCO) are often static and are rarely updated, therefore failing to capture emerging skills, which can misrepresent the true labour dynamics of particular jobs. For example, the occupational class of 'Print Journalist' has been a constant in Australian occupational statistics. Yet, the underlying skills of a 'Print Journalist' have changed dramatically in recent decades.

To overcome the above-stated limitations, in our data construction, we leveraged the Burning Glass Technologies (BGT – the job ads data source) occupational ontology together with the ANZSCO ontology. We also used the rich skill-level information from job ads that are missing from occupational employment statistics to build an encompassing journalism job ads dataset.

5.3 Data & Methods

5.3.1 Data Sources

This research used both labour demand and labour supply data to analyse journalism jobs. On the labour demand side, we used a detailed dataset of over 8 million Australian job ads, spanning from January 2012 to March 2020. These data were generously provided by Burning Glass Technologies[1] (BGT). For labour supply data, we leveraged official employment statistics [31] and salary levels [30] provided by the Australian Bureau of Statistics

[1]BGT is a leading vendor of online job ads data. `https://www.burning-glass.com/`

(ABS) over the same period. These data sources provide time series employment and salary information that have been disaggregated by gender, location, and types of employment (full-time and part-time). Further details of data sources and data construction are provided in the Supplemental Material. While there are nuances to 'journalism work' and the requirements of journalism jobs have evolved over time [215, 217, 268], this research defines journalism jobs by the official ANZSCO standards [32].

5.3.2 Skill Similarity

To analyse the underlying journalism skills within occupations, we implemented a skill similarity methodology adapted from Alabdulkareem et al. [14] and then by Dawson et al. [120] to calculate the pairwise similarities between skills from job ads.

Two skills are similar when the two are related and complementary, i.e. the two skills in a skills-pair support each other. For example, 'Journalism' and 'Editing' have a high pairwise similarity score because together they enable higher productivity for a journalist; whereas 'Journalism' and 'Oncology' have a low similarity because they are seldom required together. We measured the similarity of skill-pairs based on their co-occurrence patterns in job ads, while accounting for skill ubiquity and specialisation. To capture how journalism skills have changed over time, we measured skill similarity during calendar years.

Formally, given J as the set of job ads posted during a specific calendar year, we measured the similarity between two skills s and s' as:

$$\theta(s,s') = \frac{\sum_{j' \in J} e(j,s)e(j,s')}{max\left(\sum_{j' \in J} e(j,s), \sum_{j' \in J} e(j,s')\right)} \tag{5.1}$$

where j and j' are individuals jobs ads from the set J, and $e(s,j) \in \{0,1\}$ measures the importance of skills s for job j using theory from Trade Economics [185]. Skills s and s' are considered highly complementary if they commonly co-occur and are both 'important' for the same job ads. Finally, $\theta(s,s') \in [0,1]$, a larger value indicates that s and s' are more similar, and it reaches the maximum value when s and s' always co-occur (i.e. they never appear separately).

We build the top yearly lists of journalism skills by computing $\theta(Journalism, s)$ – i.e. the similarity between the skill 'Journalism' and each unique skill that occurs for each year from 2014-2018. The yearly top 50 skills most similar to 'Journalism' are shown in the Supplemental Material together with the full details of the θ measure.

Finally, we determined the occupations with the highest levels of skill similarity to the top journalism skills uncovered from above. We propose η, the *'Journalism Skill Intensity'*, for each standardised BGT occupation, defined as percentage of journalism skills relative to the total skill count for the job ads related to an occupation o. Formally:

$$\eta(o,\mathscr{D}) = \frac{\sum_{j\in\mathscr{O},s\in\mathscr{D}} x(j,s)}{\sum_{j\in\mathscr{O},s'\in S} x(j,s')} \tag{5.2}$$

where $\mathscr{D}$ is the set of journalism skills, and $\mathscr{O}$ is the set of job ads associated with the occupation o. This method allowed us to adaptively select occupations based on their journalism skill intensities.

5.4 Jobs Data Analysis and Results

In this section, we conducted a data-driven analysis of journalism jobs in Australia based on job ads data and official occupational statistics. First, we longitudinally examined key features of jobs data, such as employment levels, job ads posting frequency, salaries, and posting frequency growth and predictability level. We also analysed how the underlying skills of journalists had changed over time, and which skills and occupations grew in similarity to journalism.

5.4.1 Posting Frequency & Employment levels

In Australian journalism, 2012 is considered a watershed year. An estimated 1,500 journalists were made redundant, the majority of those from Australia's two largest print companies, Fairfax Media (now Nine Entertainment) and News Limited (now News Corp Australia) [347]. The severity of this industrial shock can be observed in Fig. 5.1. Against the left y-axis, the blue line shows quarterly job ads posting frequency for journalism jobs. As the graph depicts, posting frequency for journalism job ads experienced extremely low levels in 2012 until 2013, when they began to increase. The volume of vacancies increased until mid-2014, before plummeting in late-2014 to the levels last seen in 2012. From 2015, journalism job ads experienced strong growth, reaching a peak in mid-2016. Since then, journalism job ads have trended downward until the first quarter of 2020 (end of available data for job ads), albeit with volatile peaks and troughs. In summary, the data shows that journalism job ads had not been in freefall since 2012. Rather, there was erratic growth in journalism job ads until a peak in 2016, followed by erratic decline.

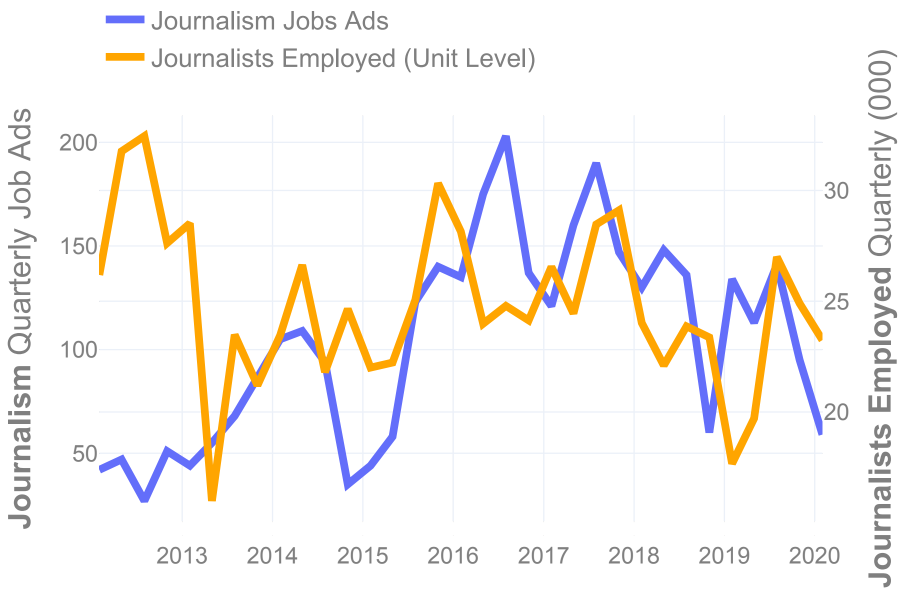


Figure 5.1: Quarterly posting frequency of journalism job ads (see Sec. Section 5.3) and employment levels of 'Journalists & Other Writers' at the ANZSCO Unit level (000's) from Jan 2012 to Mar 2020.

Similarly, employment levels underwent immense volatility from 2012 to 2013. Against the right y-axis of Fig. 5.1, the orange line shows the number of quarterly employed for 'Journalists & Other Writers' at the ANZSCO Unit level. Employment levels peaked in mid 2012, before dramatically dropping in early 2013. This is an effect of the mass journalist redundancies made in 2012, given that employment statistics are 'lagging indicators' and it takes time for labour markets to reflect changes in occupational statistics. Early 2013 marked the lowest point of journalist employment seen in this time-series. As also observed in job ads data, journalist employment levels grew until 2016-2017 and has since trended downwards, exhibiting volatile quarterly changes through to the first quarter of 2020.

COVID-19 and journalism jobs. The early effects of COVID-19 were apparent in the posting frequency of job ads in Australia. This was the case for most occupations, including journalists. Higher vacancy rates typically mean higher levels of labour demand by employers, which is a critical component of healthy labour markets. As Fig. 5.2 highlights, vacancy volumes declined for both journalism jobs and at aggregate levels in Australia. Since mid-February 2020, weekly posting frequency had decreased across all Australia job ads, as seen in Fig. 5.2a. Such a decline this early in the year is atypical. As Dawson and Rizoiu [119] show, the frequency of job ad postings follow a yearly seasonal pattern, with late February and early March typically being a period

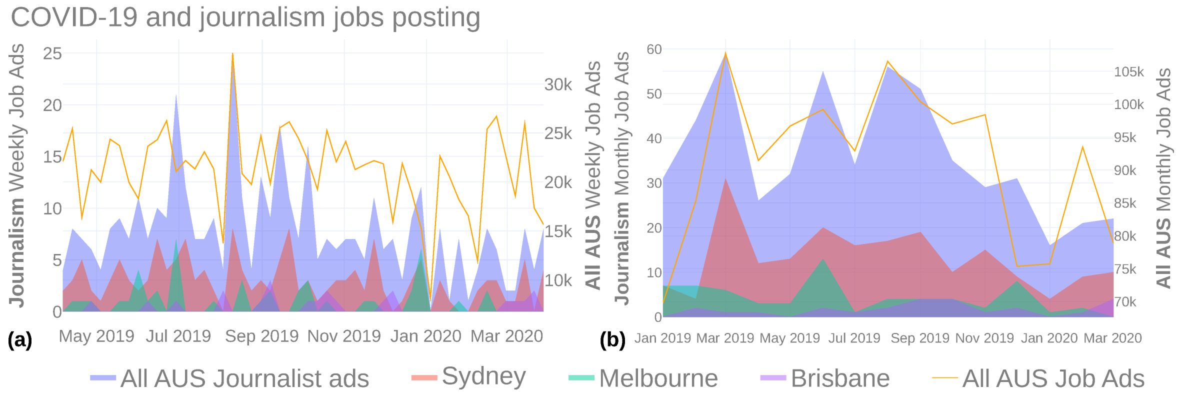


Figure 5.2: **Posting frequency for journalism jobs during the early stages of the COVID-19 crisis in Australia and its major cities**: **(a)** Weekly posting frequency volumes for journalists and all Australian job ads between April 2019 and March 2020. Both decreased as the early stages of the COVID-19 crisis hit; **(b)** Monthly posting frequency for journalists were down 63 per cent when comparing March 2019 to March 2020. This was significantly higher than all Australian job vacancies, which was down 37 per cent over the same period.

of upward trend growth. However, late February and early March 2020 coincided with the international outbreak of COVID-19. During this period, the Australian government instituted widespread quarantine and social distancing measures, which significantly constrained economic activity [76]. The impacts of these COVID-19 containment laws are starkly apparent in Fig. 5.2b. Posting frequency for journalism jobs were down 63 per cent when comparing March 2019 volumes to March 2020. This was significantly higher than the aggregate market of all Australian job ads, which was down 37 per cent over the same period. Fig. 5.2b shows that Melbourne appeared to be the city hardest hit, recording no journalism job ads in March 2020 and only 3 posts for the first quarter of 2020, even before the major lock-downs instituted for Melbourne in August 2020. Clearly the pandemic had an early and damaging effect on the journalism jobs market.

5.4.2 Salaries

We compared salaries extracted from job ads with ABS reported wage data for 'Journalists and Other Writers'[2]. Fig. 5.3 reveals two main findings regarding journalist salaries. First, according to job ads data, journalists attracted considerably lower annual wage levels (solid blue line) than the market average (dashed blue line). As of 2018, job ads

[2]ABS wage data is reported biennially, with the latest reporting year being 2018. Therefore, wage values in the 'odd' years in between the reporting periods were interpolated, calculated as the mean of the previous and the subsequent years.

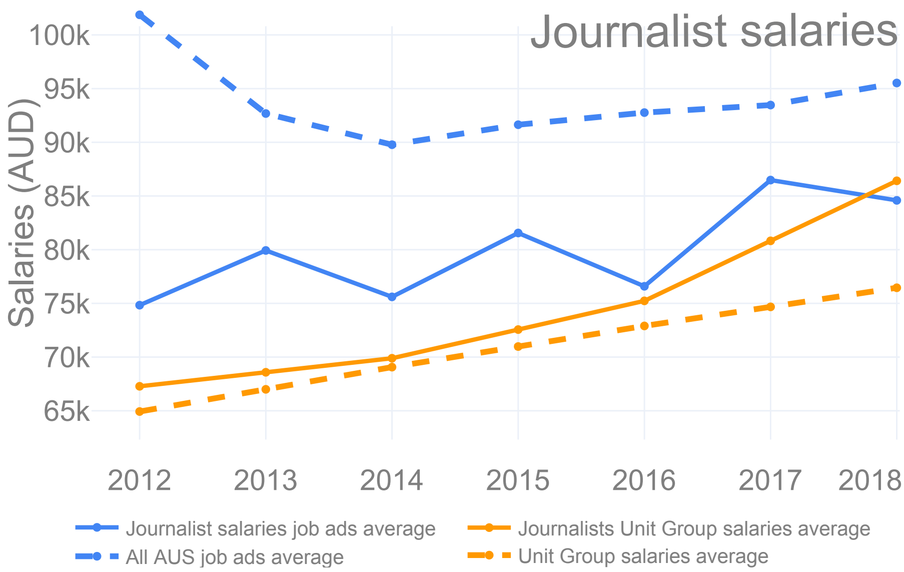


Figure 5.3: Journalist salaries (solid blue line) increased according to job ads data, but remained below market average levels (dashed blue line). However, according to ABS data, 'Journalists & Other Writers' (ANZSCO Unit level, solid orange line) earned a growing wage premium above the market average (dashed orange line).

indicated that journalists earned approximately AU$10,000 less than the market average. These findings, however, are somewhat contrary to the wage earnings data collected by the ABS [30], according to which 'Journalists and Other Writers' (solid orange line) had been earning a growing wage premium over the market average (dashed orange line) since 2014. This discrepancy can be explained by the fact that job ads data tend to over-represent occupations in the 'Professional' and 'Manager' classes [101], which typically attract higher wages. As a result, the average salary levels from job ads data (dashed blue line) were about AU$20,000 higher than average salary levels from ABS data (dashed orange line), from 2014 to 2018. However, the salary levels for journalists were very similar when comparing across the two data sources.

Fig. 5.3 yields a second observation: journalist salary levels increased in both absolute and relative terms compared to average market levels, between 2012 to 2018 in both data sources. More importantly, the relative salary growth of journalists exceeded the market averages, during the period studied.

Posting trends. We constructed an auto-regressive Machine Learning model to predict posting frequency of journalism job ads in Australia

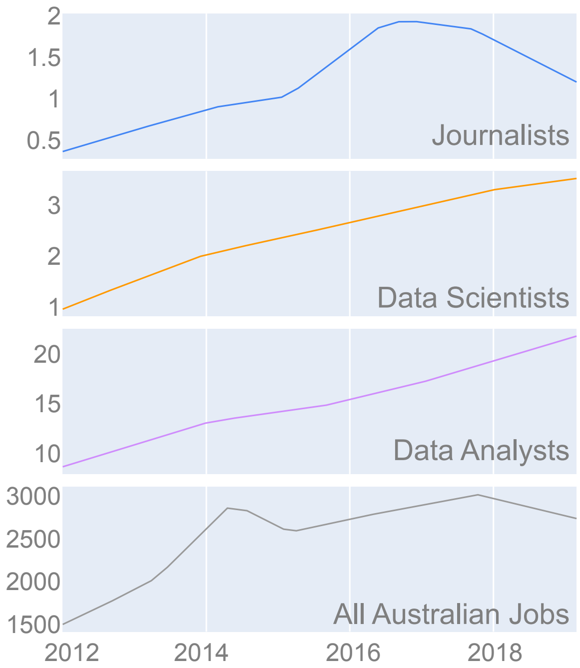


Figure 5.4: **Trend lines of posting frequency** for 'Journalists', 'Data Scientists', 'Data Analysts', and 'All Australian job ads'. Posting frequency for 'Journalists' trended downwards since 2016.

5.4.3 Trend Analysis & Predictability

Quantify labour demand volatility. When constructing Machine Learning models, it is standard procedure to use error metrics to evaluate the prediction accuracy. Volatility in posting volumes inherently lead to lowered prediction performance and higher error values. Here we use the prediction error measured using the 'Symmetric Mean Absolute Percentage Error' [220, 301] as a proxy for the volatility of labour demand for different occupations (see the technical section in the Supplemental Material for more details).

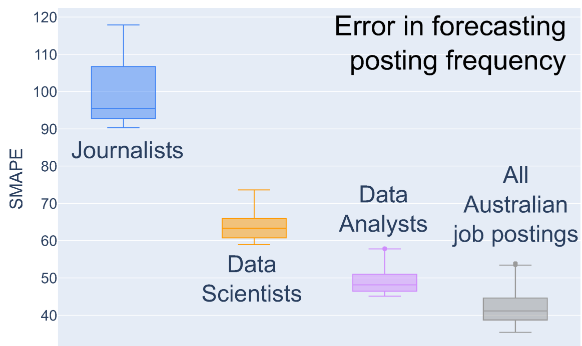
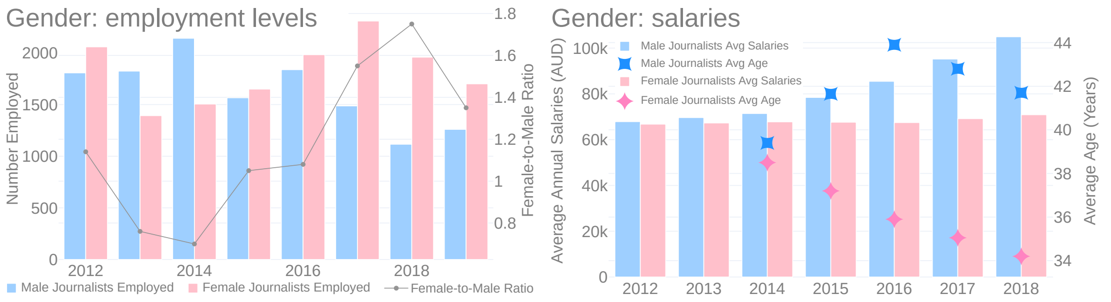


Figure 5.5: **(a)** Predictability comparison of temporal posting frequency highlighting the difficulties of predicting journalism job ads and their volatility.



((a)) ((b))

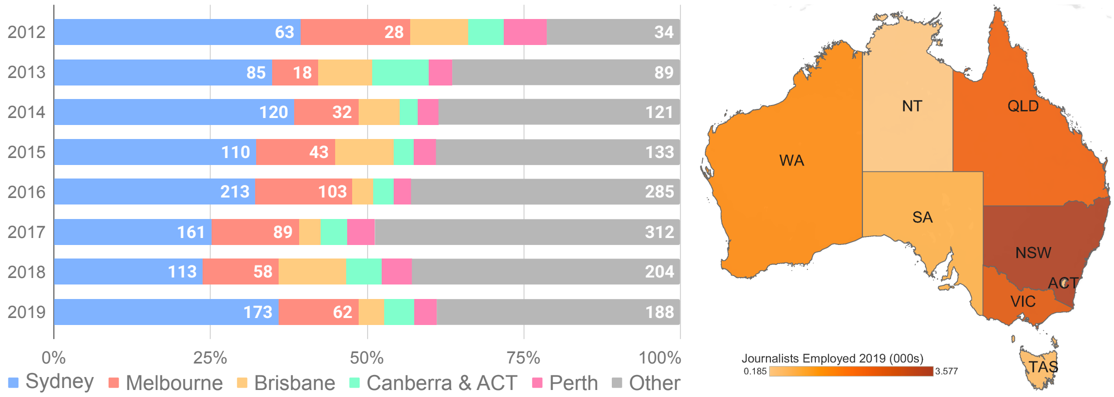
Figure 5.6: **Journalist employment levels and salaries by Gender**: **(a)** Since 2015, the employment ratio of female-to-male journalists increased; **(b)** Wage inequality increased between males and females in the 'Journalists & Other Writers' Unit group. This was at the same time that the average age of journalists decreased for females and increasing for males since 2014.

Fig. 5.5 shows the prediction performance for three occupations ('Journalists', 'Data Scientists', 'Data Analysts') and for the volume of 'All Australian job postings'. We used a sliding window approach to obtain multiple predictions (see the Supplemental Material) that we aggregated as boxplots. The higher the error score on the vertical axis, the lower the predictive abilities for that occupation. As Fig. 5.5 reveals, predicting the daily posting frequency of journalism jobs was consistently more difficult than for the other occupations, and the market as whole. 'Data Scientists', an occupation undergoing strong relative growth, is also showing a high prediction error compared to the market as a whole, indicative of experiencing a degree of volatility. However, it was not nearly commensurate to the predictive difficulties, and volatility, of journalists. This was true from 2012 to 2019, and has become worse in 2020 with the spread of COVID-19.

5.4.4 Gender

There have been growing gender differences of employed journalists in Australia [248, 249] and across the world [172]; the data presented in this research reinforces these previous findings. Fig. 5.6(a) shows that the ratio of female employed journalists increased relative to male journalists (ANZSCO Unit Level) [31]. In 2014, the female-to-male employment ratio was 0.7. In 2018, the proportion more than doubled, with almost 1.8 female journalists employed for every male journalist. It then declined in 2019 to 1.35, but this proportion was still almost double that of 2014.

Fig. 5.6(b) also shows that wage inequality between female and male journalists



((a)) ((b))

Figure 5.7: **Location of journalists in Australia**: **(a)** Posting frequency for journalism jobs decreased in major Australian cities, in relative terms; **(b)** As of 2019, the majority of journalists in Australia were employed in New South Wales, Victoria, and Queensland, respectively.

worsened from 2014 to 2018 [30]. Since 2014, the annual salaries for female journalists increased by only AU$3,000, whereas annual salaries for male journalists increased by more than AU$30,000. Male journalists thus experienced an average wage growth that was ten times greater than female journalists from 2014 to 2018.

There were also changing age demographics of employed journalists during the studied period. The markers on Fig. 5.6(b) highlight the average age of journalists by gender, per year. Male journalists were getting older, their average age increasing by two years from 2014 to 2018. Female journalists, however, were steadily getting younger. The average age for female journalists decreased by more than four years from 2014 to 2018.

5.4.5 Location

Fig. 5.7 plots the location and volume of employed journalists in Australia. Fig. 5.7(a) shows the absolute and relative number of job ads posted for each of the capital cities, and outside them, and Fig. 5.7(b) shows the location of employed journalists per state. Unsurprisingly, Sydney and Melbourne, the respective capital cities of New South Wales (NSW) and Victoria (VIC), consistently had the highest job ad posting frequencies. However, the relative share of job ad posting frequency in Australian capital cities had shrunk in later years, with Fig. 5.7(a) showing an increase outside of major cities, both in relative and absolute terms. This trend reached a peak in 2017, when less than 50 per cent of all journalist job ads were for positions inside capital cities. A small rebound

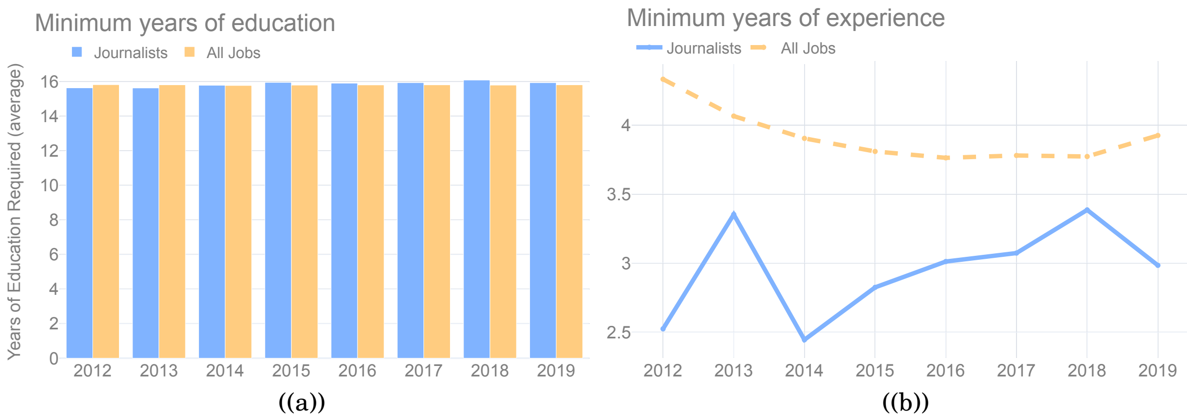


Figure 5.8: **(a)** Years of Education demanded by employers from job ads were consistent with the market average; **(b)** Years of Experience required by employers consistently remained below the market average, according to job ads. However, this gap had closed since 2014.

followed, and in 2019 Sydney commanded approximately one-third of all journalism job ads.

5.4.6 Education & Experience

Figs. 5.8(a) and 5.8(b) show respectively the number of years of formal education required for journalists, and the experience requirements (both per year, extracted from job ads data). The education requirements consistently remained at market average levels, with journalists required to possess a Bachelor-level degree (approximately 16 years of education).

By contrast, the experience requirements were more variable. Since 2012, employers required fewer years of experience from journalists than was required in the market generally. However, the gap narrowed. In 2019, employers demanded of journalists, on average, half of an additional year of experience compared to 2014. This countered the general market, where employers' demands trended downward from 2012 to 2019.

5.4.7 Employment Type

Casual and temporary work have become more commonplace in Australia [161], and we study if this is also the case for Australian journalism jobs. In Fig. 5.9 we plot the number of permanent and temporary journalism job ads, per calendar year. The number of 'Temporary' journalism jobs had increased in absolute terms since 2012, which made up the majority of all journalism ads in every year. It is noteworthy too that the share of 'Permanent' journalism vacancies had also increased since 2012. However, this trend should be interpreted with a degree of scepticism as only ~ 50 per cent of all journalism job ads specified whether the roles advertised were permanent or temporary.

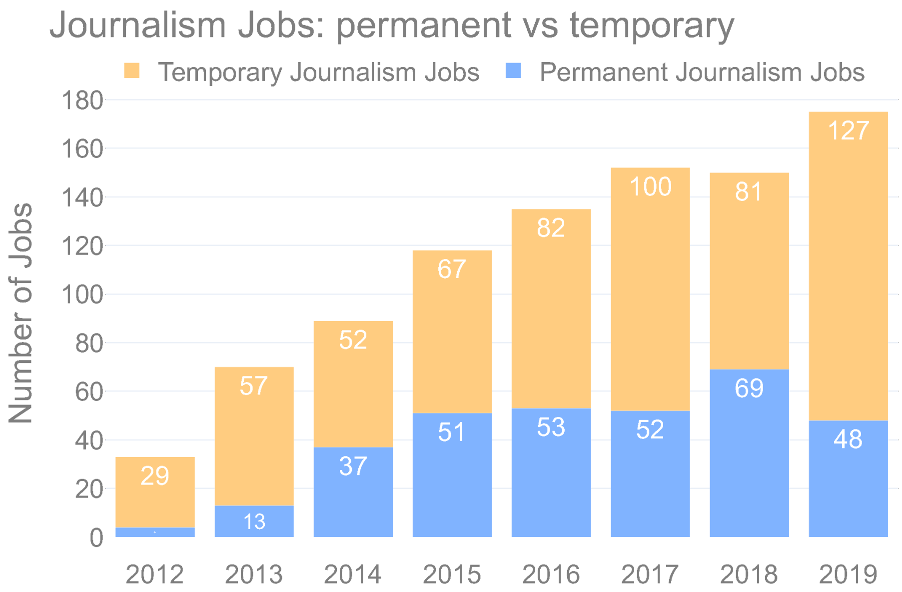


Figure 5.9: Temporary positions represented the majority of journalism job ads in Australia.

5.4.8 Journalism Skills

Growing demand for journalism skills.

Here, we analysed how the demand for some fundamental journalism skills changed over time. First, we selected three traditionally important skills to journalists that appear in job ads: (1) 'Journalism', (2) 'Editing', and (3) 'Writing'. These skills were then counted across all job ads in Australia, regardless of their occupational class. While Fig. 5.4 shows that labour demand for journalists has decreased since 2016, Fig. 5.10a presents the more nuanced story, showing that the posting frequency for each of these core journalism skills increased from 2012 to 2019, with 2018 to 2019 being the first yearly decline.

The relative rankings of these three skills also increased. For each year, we counted the posting frequency of each unique skill that appears in job ads. We then ranked these skills by posting frequency as a proxy for labour demand. Fig. 5.10b shows that the rankings of all three of these fundamental journalism skills had improved from 2012 to

2019. In other words, not only did the posting frequency of these three journalism skills increase in job ads over these eight years, but their importance relative to all other skills also increased.

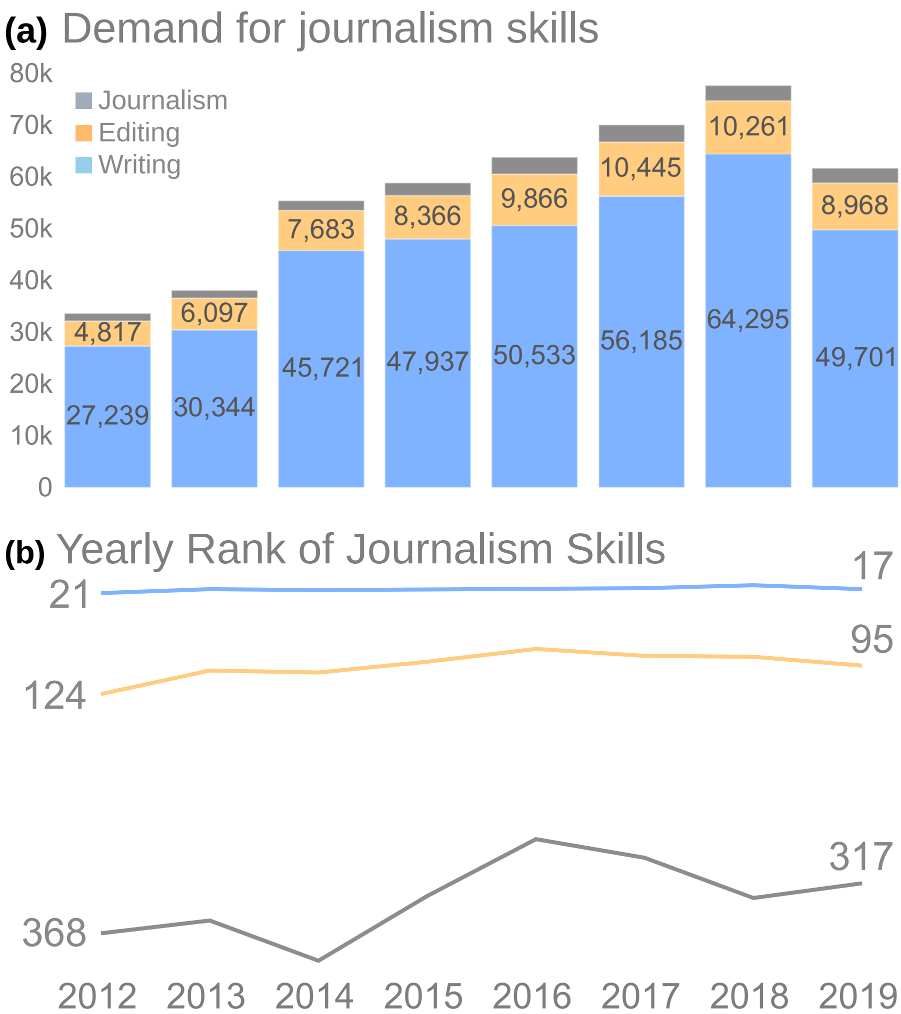


Figure 5.10: The absolute posting frequency **(a)** and relative yearly rank **(b)** of three major journalism skills increased between 2012 and 2019.

Changing importance of journalism skills. We wanted to determine whether the relative importance of the Journalism skill changed over time, using a skill similarity approach. Given the dynamics of skill requirements in job ads, skills can become increasingly more (or less) similar over time. We used the similarity measures in Eq. (5.1) to identify the skills that are becoming more relevant to being a journalist (see Section 5.3 for details, and the Supplemental Material for the top 50 skills for each from 2014 to 2018). The higher the similarity score, the more likely the skills pair will complement and support each other in a given job. Fig. 5.11(a) shows the changes in similarity scores between the skill 'Journalism' and each of the eight other top journalism skills (as per the top yearly journalism skills lists in the Supplemental Material). The greater the area covered in the radar chart, the greater the similarity score, with the blue area representing 2014 and the red area 2018.[3] Visibly in Fig. 5.11(a), 'Social Media' related skills became increasingly relevant for journalists, with the relative ratio of more traditional skills such as 'Editing' and 'Copy Writing' diminishing with respect to 'Social Media', from 2014 to 2018.

Occupations that require journalism skills. Here, we studied which occupations most required journalism skills, and their dynamics over time (according to the BGT occupational taxonomy). Given the yearly lists of top journalism skills (described in Section 5.3.2), we used Eq. (5.2) to determine the occupations with the highest intensities of journalism skills, for each year from 2014 to 2018. Intuitively, this allows us adaptively to identify occupations that become more or less similar to 'Journalism', based on their underlying skill usage. It also provides a means to assess likely transitions between occupations, as workers are more likely to transition to occupations where the underlying skill requirements are similar [57]. Higher similarity lowers the barriers to entry from one occupation to another.

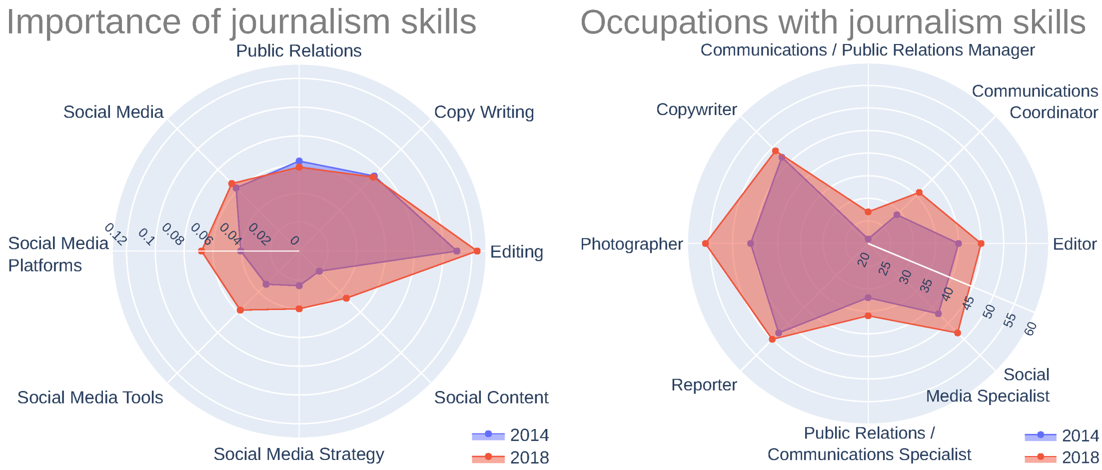
Fig. 5.11(b) highlights eight top occupations and their journalism skill intensity scores for 2014 and 2018. 'Reporter', 'Editor', and 'Copywriter' cover the highest percentage of journalism jobs in the dataset, respectively. While the journalism skill intensities of these occupations were relatively high in 2018, their growth since 2014 was relatively low. In comparison, 'Photography', 'Communications', 'Social Media', and 'Public Relations' experienced higher journalism skill intensity growth from 2014 to 2018. This provides insights as to where workers with journalism skills might have found employment outside of journalism.

5.5 Discussion

5.5.1 Volatility of journalism jobs

Drawn from job ads and employment statistics, our findings reveal the highly volatile nature of the journalism industry. Compared to other occupations and the aggregate labour market, journalism experiences dramatic fluctuations that are unpredictable and irregular (see Fig. 5.5). The data also confirmed that journalism is an industry in crisis,

[3]At the time of analysis, 2018 was the final full year available of access to the required skills-level data.



((a)) ((b))

Figure 5.11: **Skill and occupational similarity analyses**: **(a)** The changing similarity (or relative importance) of specific skills compared to the skill 'Journalism'; **(b)** Eight occupations that had the highest similarity to the 'Top Yearly Journalism Skills'.

worsened in the early stages of COVID-19. However, the data also reveals surprises, including that the number of journalism jobs ads and employment levels *increased* from 2012 until 2016. Since then, though, journalism jobs in Australia declined.

The volatility of journalism jobs in Australia was clearly apparent in Section 5.4.1. Posting frequency of job ads ranged from near zero levels in 2012 and 2014 to more than 200 posts per quarter in 2016. These violent swings are also apparent in the quarterly employment statistics of 'Journalists and Other Writers'. Following the mass redundancies of 2012, employment levels plummeted, reaching their lowest levels in 2013. They then increased before falling again into the beginning of the COVID-19 pandemic. However, the data confirms that volatility of employment has been a constant for journalism, and that this has worsened during COVID-19.

Fig. 5.5 reveals this extreme volatility. The error metrics from the Machine Learning model used to predict daily posting frequencies of job ads (as detailed in Section 5.4.3) highlight the difficulties of making predictions about journalism employment. This lack of predictability is indicative of volatility. The higher the error scores for a given occupation, the higher the likelihood that the occupation is experiencing significant disruption. This becomes apparent when we compare journalism to other occupations. For example, the volatility of 'Journalists' dwarfs that of 'Data Scientists', an occupation experiencing significant demand and volatility in Australia Fig. 5.5.

The volatility of journalism jobs was further revealed by a time series analysis of journalism compared to other occupations (Fig. 5.4), a gender-based analysis (Fig. 5.6), a geographical analysis (Fig. 5.7) and an analysis of the temporary nature of journalism jobs (Fig. 5.9).

What is indisputably clear is that the advertising market for news and journalism collapsed, and, at the time of writing, continues to collapse [6, 133, 305]. Meanwhile, consumers have tended to show an unwillingness to pay for digital journalistic content:

in 2019, Australian news consumers admitted they would much would rather subscribe to a video streaming service such as Netflix (34 per cent), than pay for online news (9 per cent) [145]. Admittedly, during COVID-19 some subscription rates have risen [137]. Clearly, however, the Internet has detonated the advertising model that once sustained journalism, and simultaneously re-adjusted consumer expectations on the monetary value of journalism content. The fact that journalism is struggling is confirmed in several ways by the data, including by the unpredictability of job ads posting frequency and the clear shifts in employment levels, as shown in Fig. 5.1. To say that journalism has been disrupted is an understatement.

Volatility exacerbated by COVID-19. In a fragmenting news ecosystem, consumer demand for news and journalism is difficult to quantify. The *Digital News Report: Australia 2019* has found that many consumers are disengaging, with the proportion of Australians avoiding news increasing from 57 per cent in 2017 to 62 per cent in 2019 [145]. Demand for 'quality' and 'public interest' journalism is even harder to quantify, given ongoing debates as to what exactly constitutes 'quality' and 'public interest' [337]. Nonetheless, demand for journalism has surged dramatically since the outbreak of COVID-19.

The irony of the coronavirus pandemic is that even as it has been killing off journalism jobs, it has also created a heightened demand for, and appreciation of, journalism among the general public. As news analyst Ken Doctor [133] wrote of the US situation in late March, 'The amount of time Americans spend with journalists' work and their willingness to pay for it have both spiked, higher than at any point since Election 2016, maybe before ... [but] how many journalists will still have jobs once the initial virus panic subsides?'. In the UK in March, *The Guardian* received 2.17 billion page views, an increase of more than 750 million above its previous record, set in October 2019 [60].

Since the outbreak of COVID-19, the volatility of the journalism jobs market has worsened dramatically. We noted above that in May News Corp ended the print run of more than 100 newspapers nationally In April, Australian Community Newspapers, which publishes 170 community titles, said it was suspending publication of some of its non-daily newspapers; as a result, four printing presses were closed and an unspecified number of staff were stood down [231]. Also in April, the federal government announced a AU$50million package to support public interest journalism across TV, newspapers and radio in regional and remote Australia [179]. And on April 20, the Australian government announced that digital platforms including Google and Facebook would be forced to pay for content as the Internet advertising business would be overhauled to help local publishers survive the economic fallout of the coronavirus crisis [115]. The scheme, which would involve a mandatory code imposed on digital giants, would potentially set a global precedent. The combined and ongoing impact on journalism jobs of these sudden, cumulative developments are hard to predict, but will no doubt be profound.

5.5.2 Gender Wage Gap

At first glance, the data seems to suggest that gender equity is finally arriving in Australia for journalism – an industry that has traditionally been male-dominated – as more women than men are employed. As the data shows, in 2014 there were 0.7 female journalists employed for every male Journalist, but by 2018 the proportion of female-to-male employment more than doubled, with almost 1.8 female journalists employed for every male Journalist. It then declined in 2019 to 1.35, a proportion still almost double that of 2014.

However, further detail reveals that equity remained elusive. Specifically, wage inequality worsened. Since 2014, annual salaries for female journalists increased by AU$3,000, compared with an increase for male journalists of over AU$30,000 over the same period. From 2014 to 2018, average wage growth for Male journalists was more than ten times greater than for female journalists. Meanwhile, the average male Journalist was getting older, while the average female Journalist was getting younger. In 2014, the average age for a Journalist, whether male or female, was roughly the same: late 30s. By 2018, the average age for a male journalist was 42, whereas for a female journalist it was 34. These results support previous findings on the changing demographic characteristics of journalists in Australia. In a survey of female journalists in Australia, North [249] found gendered divisions of tasks associated with reporting, where the majority of female reporters were assigned ‘soft-news’ areas, such as arts, education, and health. These gender and age inequities for journalists were also present in other countries [172]. The wage and age discrepancies between female and male journalists observed in the employment statistics are consistent with the surveyed experiences of female journalists in Australia by North [249].

The potential impacts of this worsening disparity are concerning. It is possible that senior positions responsible for major editorial decisions were increasingly being dominated by men, whereas junior roles were being filled by women who are younger and worse-paid.

Further research is needed into related issues of the industry's composition, including, for instance, the ethnicity of journalists. A vast body of literature exists regarding the importance of diversity in news [91, 290]. Further work is needed into diversity (and its various sub-categories), and what effect diversity has, for instance, on the proportion of people who are actively avoiding the news.

5.5.3 Location

As discussed above, the sustained pressures on regional and local journalism have led to a worrying growth of 'news deserts' in countries including Australia and the US. This trend alarmingly accelerated in the early stages of COVID-19, leaving many areas without any regional or local news coverage. For instance, as of October 2020, 'Public Interest Journalism Initiative' had documented a net decline of 124 newsrooms from January 2019 [278]. Hence, we might assume that journalism jobs in regional and local areas were been drying up, and that an ever-increasing proportion of journalism jobs were in urban centres.

The data, however, were not so clear until the end of 2019. As Fig. 5.7(a) shows, in 2012 fewer than a quarter of Australia's journalism job ads were for jobs outside Sydney, Melbourne, Brisbane, Canberra and the ACT or Perth. In every subsequent year, the proportion of job ads for journalism positions outside these urban centres was considerably higher. The peak came in 2017, when nearly half of all job ads were for positions outside the major cities. Does this suggest that in 2017 there were as many jobs for journalists in the regions as in the centres? Surely not. The explanation, we suggest, lies in various factors. These include that regional journalism jobs are hard to fill, perhaps because they offer relatively low salaries, and are hence re-advertised. It is also possible that there is a high turnover for some regional positions. In short, the job ads data may simply be an indication that the journalism industry is even more volatile in the regions than in major urban centres.

Research consistently and emphatically reveals that regional and local journalism have been suffering, with an increasingly bleak prognosis of cuts and closures [5, 6, 133]. While the data shows a surprisingly high proportion of journalism job ads for positions outside the main metropolitan centres, this cannot be taken to suggest that journalism is holding steady in these areas.

5.5.4 Evolving journalism skills

Skills are the building blocks of jobs and standardised occupations. In this regard, occupations can be characterised as 'sets of skills'. Intuitively, skills that are similar can be interpreted as complementary when they are paired together or relatively easy to acquire when one skill is already possessed.

This intuition provides insight into how journalism skills are evolving and where journalists might be finding alternate career paths. As Fig. 5.1 shows, both the demand for and supply of journalists have been declining in Australia since 2016. Therefore, a growing number of former journalists, who presumably possess an assortment of journalism skills, needed to transition between occupations to find new work. There are, however, significant transition costs moving between jobs [57, 68]. These costs can come in the form of education, training, physically moving for new employment and other barriers. To reduce the friction of these transition costs, workers tend to leverage their extant skills, in concert with acquiring new skills, to make career transitions.

As seen in Fig. 5.11(a), the skill 'Journalism' became more similar to 'Social Media' and more 'generalist' communications skills. After applying the *Skill Intensity* formula from Eq. (5.2), we identified the top occupations with highest intensities of journalism skills from 2014-2018 – these normalised measures from Eq. (5.1) and Eq. (5.2) take into account newly emerging and redundant skills. The Fig. 5.11(b) chart reinforces that top journalism skills were becoming more important to other occupations, such as 'Photographers', 'Social Media Strategists', 'Public Relations Professionals', and 'Communications Specialists'. While it is certainly possible that journalism tasks were being performed in these different occupations, it nonetheless highlights the changing nature of journalism work and the occupations where journalism skills were of growing importance.

From the data, we suggest, three conclusions can be drawn, which supports previous research [216, 217, 268, 345]. First, to be hired, journalists are required to have a wider array of skills, such as photography and social media aptitude. Second, jobs requiring journalism skills were increasingly occupations in social media, generalist communications, and public relations rather than in reporting and editing. And third, we see hints as to where onetime journalists are finding alternate career paths. As employment conditions progressively worsen, journalists are seemingly pursuing new careers in the occupational areas seen in Fig. 5.11(b), such as photography or public relations.

At a time of great uncertainty, with employment prospects deteriorating, it is no wonder that journalists look beyond traditional journalism for their futures. For society, however, the implications are significant. In this time of economic instability and polarising politics, the people who possess the journalism skills required to keep the public informed and hold leaders to account are, in many cases, employing their talents elsewhere. This places enormous strain on the health and quality of journalism in Australia.

5.6 Conclusion

The data reveals a contradiction: demand for journalism skills increased at the same time that demand and employment for journalists declined. Indeed, this is one of several contradictions in a volatile industry. For an increasing number of news media organisations, a sustainable business model remains elusive.

Our findings give a clearer outline of the problem. Unfortunately, the solutions remain less clear. Quality journalism is expensive. Good reporting is often slow and laborious, fixed to the unfolding story. What is required of quality journalism is, therefore, at odds with the prevailing employment conditions.

This paper highlights the stresses experienced by journalism in Australia by analysing jobs data. We observed the volatility and downward trajectory of the occupation both in job ads and employment statistics. These unfavourable employment conditions were worsened by the unfolding COVID-19 crisis. Our longitudinal analysis also yields important findings regarding gender inequity. While women represented a greater share of employed journalists, they earned less, and the wage gap grew.

Further, this paper also identified top journalism skills. Adopting a data-driven method, we described which skills are most similar to 'Journalism'. We then used these yearly skill sets to adaptively select similar occupations. This enabled us to quantitatively show that the skill demands of journalists became increasingly similar to those of 'Social Media Strategists', 'Public Relations Professionals', 'Communications Specialists', and others. This suggests where people with journalism skills were likely finding alternate career paths, but also raises a related concern. On the face of it, the journalism jobs data we have analysed does not look so bad after all. On reflection, however, it suggests that the thinning ranks of 'journalism' are populated by fewer journalists, and more public relations specialists.

Future research could compare these results to other labour markets to assess the validity of these findings. For example, the skill similarity methodology could be applied in other labour markets to compare the resulting top journalism skills in different locations. Additionally, labour demand analyses could be conducted on occupations most similar to journalists to better understand the incentives to transition to other vocations. Further work could also examine the implications of changing journalism skill demands for journalism schools. This research demonstrated that not only have the skills demanded of journalists evolved, but the occupations that require journalism skills have broadened. The extent to which journalism schools are adequately preparing its students for the quickly changing labour demands of journalists is a rich area of inquiry.

The results from this research both reinforce the well-documented difficulties of journalism in Australia and provide granular details that isolate and reveal these challenges. The hope is that these analytical methods and insights can contribute to the health and well-being of the Fourth Estate, and hence to the health and well-being of society.

5.7 Supplementary Materials

These Supplementary Materials is accompanying the submission *Layoffs, Inequity and COVID-19: A Longitudinal Study of the Journalism Jobs Crisis in Australia from 2012 to 2020*. The information in these Supplementary Materials complements the submission, and it is presented here for completeness reasons. It is not required for understanding the main paper, nor for reproducing the results.

Here, we describe the data sources we used to analyse journalism jobs. We also outline the skill similarity methodology that enables us to construct temporal (yearly) sets of top journalism skills. Lastly, we describe how these temporal sets of top journalism skills then allow us to adaptively identify occupations that are 'most similar' to journalism, at the granular skill level.

5.7.1 Data Sources

Journalism job ads. This research draws on more than 8 million Australian online job ads from 2012-01-01 until 2019-02-28, courtesy of data provided by Burning Glass Technologies[4] (BGT). BGT also granted access to the aggregated job ads data from 2019-03-01 to 2020-03-31, allowing us to address the early impacts of the unfolding coronavirus pandemic (COVID-19) on journalism jobs in Australia. BGT collected the job ads data via web scraping and systematically processed it into structured formats. The dataset consists of detailed information on individual job ads, such as location, salary, employer, educational requirements, experience demands, and more. The skill requirements have also been extracted (totalling $> 11{,}000$ unique skills) and each job ad is classified into its relevant occupational and industry classes. There are two occupational ontologies in the job ads dataset. The first is ANZSCO, which is the official occupational classification standard in Australia and New Zealand. The other is the BGT occupational ontology, which has been developed due to shortcomings of official occupational standards (as described in *Related Work & Background*).

To ensure selection accuracy, we instituted the following search query conditions over the dataset:

1. All job ads with ANZSCO Occupation labels of 'Newspaper or Periodical Editor', 'Print Journalist', 'Radio Journalist', 'Television Journalist', and 'Journalists and Other Writers nec' (where 'nec' stands for 'not elsewhere classified').

[4]BGT is a leading vendor of online job ads data. `https://www.burning-glass.com/`

2. OR All job ads with the BGT Occupation label of 'Journalist / Reporter' and 'Editor' (the two primary BGT occupational classes for journalists);

3. OR All job ads with the 'Journalist', 'News', or 'Editor' in any part of the job title.

After manually reviewing the returned job ad features for accuracy, the selection process resulted in a sample of 3,231 Australian journalism job ads from 2012-01-01 until 2019-02-28. We used the same search query and approach for the 2019-03-01 to 2020-03-31 period to supplement this sample. This returned 467 journalism job ads, amounting to a total of 3,698 journalism job ads from 2012-01-01 to 2020-03-31. The job ads during the period are observed aggregated daily, with limited skill level details. However, much of the analysis that follows requires access to the features within individual job ads, so only Fig. 2 leverages the 2020 data.

Further details on job ads data. It is estimated that approximately 60% of Australian job ads are posted online [126], which grew quickly in the early 2000's before plateauing in recent years [101]. At aggregate levels, online job advertisements (ads) provide valuable indicators of relative labour demands. This includes demand features, such as salaries, educational requirements, years of experience, and, most importantly, skill-level information. Here, a distinction must be made between skills, knowledge, abilities, and occupations. 'Skills' are the proficiencies developed through training and/or experience [263]; 'knowledge' is the theoretical and/or practical understanding of an area; 'ability' is the competency to achieve a task [157]; and 'occupations' are standardised jobs that are the amalgamation of skills, knowledge, and abilities used by an individual to perform a set of tasks that are required by their vocation. Throughout this paper, the term 'skill' will incorporate 'knowledge' and 'ability'. Skills, in this sense, are the constituent elements that workers use to perform tasks, which ultimately define jobs and occupations. While it is possible that the ways skills are described could evolve over time, it is unlikely that their meanings materially changed over the nine year period analysed in this research.

Advantages of job ads data. Understanding how the composition of skill sets evolve within an occupation is essential to understanding trends in that occupation. However, occupational data rarely captures skill-level data. Most often, official occupational standards are static, rarely updated classifications, which fail to capture the changing skill demands of occupations, or to detect the creation of new types of jobs.

Example of journalism job ad titles. The table below illustrates a random sample of job ad titles classified as journalists in the BGT dataset.

Table 5.1: Random sample of journalism job ad titles

Job Ad Title
News And Features Journalist - The Courier
Journalist/Copywriter
Senior Finance Journalist/Content Manager
ABC Rural Reporter
Rural Reporter, ABC Local Radio
Newspaper Journalist
Journalist Radio News - Illawarra
News Writer
Editor, News
Senior Digital News Editor - Leading Digital Media Company

Journalist employment statistics. Employment data (labour supply) were collected from the 'Quarterly Detailed Labour Force' statistics by the ABS [31]. These employment data are organised into standardised occupations called the Australia and New Zealand Standard Classification of Occupations (ANZSCO). ANZSCO provides a basis for the standardised collection, analysis and dissemination of occupational data for Australia and New Zealand. The structure of ANZSCO has five hierarchical levels - major group, sub-major group, minor group, unit group and occupation. The categories at the most detailed level of the classification are termed 'occupations'.

A shortcoming, however, is that the lowest level of occupational employment data available by the ABS is at the 4-digit Unit level, which is one hierarchical level above specific occupations. As our research is focused on the employment Unit class of 'Journalists and Other Writers', all ABS employment statistics cited in this research include the following occupations: 'Copywriter', 'Newspaper or Periodical Editor', 'Print Journalist', 'Radio Journalist', 'Technical Writer', 'Television Journalist', and 'Journalists and Other Writers nec'. While the inclusion of the 'Copywriter' and 'Technical Writer' occupations in these statistics could distort results pertaining to 'Journalists' to an extent, we consider this impact to be limited in scope. As we describe in *Jobs Data Analysis and Results*, the employment statistics highlight important trends in journalism occupations, which are confirmed by findings from the job ads data.

Another shortcoming of employment statistics is their 'lagging' nature. The inertia of labour markets means that it takes time for changes to materialise in employment statistics. Additionally, the official reporting of employment statistics takes time. Employment statistics are often published several months or years after the reported period.

As a result, these 'lagging' characteristics are not available for the most recent periods in our work (such as for the second half of 2019 and later.)

5.7.2 Skill Similarity

In this section, we detail the methodology previously employed in [14, 120] to dynamically measure skill similarity. Here, we present the building blocks for this method, applying it for journalism related skills and occupations.

Intuition. Two skills are similar when the two are related and complementary, i.e. the two skills in a skills-pair support each other. For example, 'Journalism' and 'Editing' have a high pairwise similarity score because together they enable higher productivity for the worker, and because the difficulty to acquire either skill when one is already possessed by a worker is relatively low.

Our goal, therefore, is to calculate the similarity of each unique skill relative to every other unique skill in the dataset. Such a measure allows us to identify which skills have the highest pairwise similarities to a specific skill or set of skills. We also want to identify how skill similarity evolves over time. To achieve this, we have instituted a temporal split of a calendar year. This enables us to assess yearly changes to the underlying skill demands of journalism jobs.

The Revealed Comparative Advantage of a skill. We implement a data-driven methodology to measure the pairwise similarity between pairs of skills that co-occur in job ads. One difficulty we encounter is that some skills are ubiquitous, occurring across many job ads and occupations. We address this issue by using the *Revealed Comparative Advantage* (RCA), which maximises the amount of skill-level information obtained from each job ad, while minimising the biases introduced by over-expressed skills in job ads. Formally, RCA measures the relevance of a skill s for a particular job ad j as:

$$RCA(j,s) = \frac{x(j,s) / \sum_{s' \in \mathscr{S}} x(j,s')}{\sum_{j' \in J} x(j',s) / \sum_{j' \in \mathscr{J}, s' \in \mathscr{S}} x(j',s')} \tag{5.3}$$

where $x(j,s) = 1$ when the skill s is required for job j, and $x(j,s) = 0$ otherwise; $\mathscr{S}$ is the set of all distinct skills, and $\mathscr{J}$ is the set of all job ads in our dataset. $RCA(j,s) \in \left[0, \sum_{j' \in J, s' \in S} x(j',s')\right], \forall j,s$, and the higher $RCA(j,s)$ the higher is the comparative advantage that s is considered to have for j. Visibly, $RCA(j,s)$ decreases when the skill s is more ubiquitous (i.e. when $\sum_{j' \in J} x(j',s)$ increases), or when many other skills are required for the job j (i.e. when $\sum_{s' \in S} x(j,s')$ increases). RCA provides a method to measure the importance of a skill in a job ad, relative to the total share of demand for that skill in all job ads. It has been applied across a range of disciplines, such as trade economics [185] [330], identifying key industries in nations [307], and detecting the labour polarisation of workplace skills [14].

Measure skill similarity. The next step is measuring the complementarity of skill-pairs that co-occur in job ads. First, we compute the 'effective use of skills' $e(j,s)$ defined as $e(j,s) = 1$ when $RCA(j,s) > 1$ and $e(j,s) = 0$ otherwise. Finally, we compute the skill complementarity (denoted θ) as the minimum of the conditional probabilities of a skills-pair being effectively used within the same job ad. Skills s and s' are considered as highly complementary if they tend to commonly co-occur within individual job ads, for whatever reason. Formally:

$$\theta(s,s') = \frac{\sum_{j' \in J} e(j,s).e(j,s')}{max\left(\sum_{j' \in J} e(j,s), \sum_{j' \in J} e(j,s')\right)} \tag{5.4}$$

Note that $\theta(s,s') \in [0,1]$, a larger value indicates that s and s' are more similar, and it reaches the maximum value when s and s' always co-occur (i.e. they never appear separately).

Top journalism skills. Following the procedure outlined in Eq. (5.4) for building sets of highly complementary skills, we use the θ function together with 'Journalism' as the 'seed' skill to create top yearly lists of journalism skills. More precisely, we compute $\theta(Journalism,s)$ – i.e. the similarity between the skill 'Journalism' and each unique skill that occurs during a given year. Skills on each yearly list are ordered by their descending pairwise skill similarity scores. When inspecting the yearly skill lists, we make two observations. First, the skills in 2012 and 2013 appear of notably lower quality than from 2014 onward. We posit that this has to do with imperfect skills extraction methods during the early years of the BGT dataset. As a result, we decided to measure the top yearly journalism skill sets from 2014 to 2018 (the last available full year of data for which we had access). [5] Second, we decided to retain only the top 50 skills on each yearly list. Through qualitative analysis, we determined that this threshold of 50 is both sufficiently exclusive for defining journalism skills and reasonably inclusive for detecting the evolution of new, emerging skills in journalism. The purpose of these

[5] We did not notice a deterioration of quality regarding other features, such as salaries, education, experience etc. Therefore, these 2012 and 2013 will be used for parts of the analysis.

top journalism skills lists is to capture journalism labour trends; it is not intended to represent a complete taxonomy of journalism skills. The yearly lists of top journalism skills, and their similarity scores, can be observed in Section 5.7.5.

Compute journalism skill intensity. For the occupational similarity analysis in Sec. *Journalism Skills*, we decided to use the BGT occupational ontology as opposed to ANZSCO. This is because the BGT occupational classes appear more reflective of current job titles. For example, a job title advertised for a 'Social Media Manager' is classified by BGT as a 'Social Media Strategist / Specialist'. Whereas the same job title would be classified by ANZSCO as an 'Advertising Specialist' or 'Marketing Specialist'.

5.7.3 Trend Analysis & Predictability

We use the Prophet time-series forecasting tool developed by Facebook Research [319]. Prophet is an auto-regressive tool that fits non-linear time-series trends with the effects from daily, weekly, and yearly seasonality, and also holidays. The main model components are represented in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{5.5}$$

where $g(t)$ refers to the trend function that models non-periodic changes over time; $s(t)$ represents periodic changes, such as seasonality; $h(t)$ denotes holiday effects; and ϵ_t is the error term and represents all other idiosyncratic changes.

5.7.4 Quantify Labour Demand Volatility

We evaluate the forecasting performance using a temporal holdout setup. That is, we split the available time-series into a training part (the first part of the sequence) and a testing part (the latter part of the sequence). We train the Prophet model on the training part, and we generate job ad posting forecasts by "running time forward" in Eq. (5.5) for time t in the testing period. Finally, we measure the accuracy of the forecast against the observed posting volumes using the Symmetric Mean Absolute Percentage Error (SMAPE) [220, 301]. SMAPE is formally defined as:

$$SMAPE(A_t, F_t) = \frac{200}{T} \sum_{t=1}^{T} \frac{|F_t - A_t|}{(|A_t| + |F_t|)} \tag{5.6}$$

where A_t denotes the actual value of jobs posted on day t, and F_t is the predicted value of job ads on day t. SMAPE ranges from 0 to 200, with 0 indicating a perfect prediction and 200 the largest possible error. When actual and predicted values are both 0, we define SMAPE to be 0. We selected SMAPE as an alternative to the more widely used MAPE because it is (1) scale-independent and (2) robust to actual or predicted zero values. To evaluate the uncertainty of the forecast, we adopt a 'sliding window' approach. This consists of using a constant number of training days (here $1,186$ days) to train the model, and we test the forecasting performance on the next 365 days. We then shift both the training and the testing periods right by one day, and the process is repeated. Consequently, we train and test the model 365 times, and we obtain 365 SMAPE performance values.

5.7.5 Top Journalism Skills by Year

Top journalism skills calculated by skill similarity methodology in Sec. *Skill Similarity*.

Table 5.2: Top journalism skills calculated by skill similarity methodology in *Skill Similarity*

Rank	2014	2015	2016	2017	2018
1	Journalism	Journalism	Journalism	Journalism	Journalism
2	Editing	Editing	Editing	Editing	Editing
3	Media Relations	Media Relations	Copy Writing	Content Management	Content Management
4	Corporate Communications	Copy Writing	Media Relations	Social Media	Media Relations
5	Copy Writing	Content Management	Content Management	Copy Writing	Copy Writing
6	Content Management	Copywriting	Social Media	Media Relations	Social Media Platforms
7	Public Relations	Social Media	Social Media Platforms	Corporate Communications	Social Media
8	Social Media	Public Relations	Copywriting	Social Media Platforms	Content Development
9	Content Management Systems (CMS)	Social Media Platforms	Corporate Communications	Content Development	Corporate Communications
10	Multimedia	Corporate Communications	Public Relations	Social Content	Public Relations
11	Copywriting	Content Development	Content Development	Public Relations	Social Media Tools
12	Content Development	Content Management Systems (CMS)	Social Media Tools	Copywriting	Copywriting
13	Strategic Communications	Strategic Communications	Digital Marketing	Facebook	Content Management Systems (CMS)
14	Facebook	Social Media Tools	Online Marketing	Strategic Communications	Social Content
15	Social Media Platforms	Multimedia	Multimedia	Social Media Tools	Strategic Communications
16	Marketing Communications	Facebook	Strategic Communications	Marketing Communications	Social Media Strategy
17	Media Coverage	Marketing Communications	Market Research	Content Management Systems (CMS)	Content Marketing
18	Publicity	Social Content	Marketing Communications	Multimedia	Facebook
19	Proofreading	Digital Communications	Content Management Systems (CMS)	Proofreading	Digital Communications
20	Social Media Tools	Publicity	Writing	Content Marketing	Media Coverage
21	Digital Communications	Media Production	Content Marketing	Digital Journalism	Publicity
22	Crisis Management	Social Media Strategy	Photography	Digital Communications	Proofreading
23	Adobe Photoshop	Communications Programmes	Instagram	Publicity	Multimedia
24	Communications Programmes	Media Coverage	Publicity	Media Coverage	Instagram
25	Digital Journalism	Internal Communications	Digital Communications	Digital Marketing	Video Production
26	Community Relations	Content Marketing	Media Coverage	Writing	Marketing Communications
27	Photography	Proofreading	Social Content	Video Production	Adobe Photoshop
28	Social Media Strategy	Writing	Social Media Strategy	Graphic Design	Content Curation
29	Graphic Design	Adobe Photoshop	Media Production	Media Production	Video Editing
30	Youtube	Brand Awareness Generation	Proofreading	Communications Programmes	Adobe Indesign
31	Media Strategy	Adobe Indesign	Facebook	Instagram	Adobe Creative Suite
32	Brand Management	Marketing Materials	Event Planning	Social Media Strategy	Adobe Acrobat
33	Web Content Management	Digital Marketing	Adobe Photoshop	Video Editing	Brand Awareness Generation
34	Adobe Indesign	Video Editing	Meeting Deadlines	Adobe Photoshop	Adobe Illustrator
35	Social Content	Adobe Creative Suite	Self-Starter	Self-Starter	Google Analytics
36	Marketing Materials	Adobe Acrobat	Marketing	Breaking News Coverage	Press Releases
37	Event Planning	Graphic Design	Creativity	Creativity	LinkedIn
38	Digital Marketing	Video Production	Adobe Indesign	Adobe Illustrator	Digital Marketing
39	Writing	Instagram	Adobe Creative Suite	Event Planning	Digital Journalism
40	Instagram	LinkedIn	Adobe Illustrator	Adobe Indesign	Media Production
41	Online Research	Media Strategy	Community Relations	Adobe Creative Suite	Communications Programmes
42	Adobe Acrobat	Photography	Adobe Acrobat	Adobe Acrobat	Crisis Management
43	LinkedIn	PR Agency	Press Releases	Promotional Materials	Media Strategy
44	Google Analytics	Meeting Deadlines	Internal Communications	Photography	Writing
45	Video Editing	Digital Journalism	Campaign Management	Content Curation	Photography
46	Website Production	Event Planning	Creative Writing	Marketing	Blog Posts
47	Proofing	Google Analytics	Video Production	Meeting Deadlines	Internal Communications
48	Video Production	Media Campaigning	Blog Posts	Google Analytics	Event Planning
49	Media Planning	Press Releases	Crisis Management	Media Strategy	Creative Problem Solving
50	Campaign Management	Crisis Management	Youtube	Business-to-Business	Creativity

CHAPTER 6

Paper 4 - Skill-driven Recommendations for Job Transition Pathways

Preamble

This paper proposes a novel method to measure the relative similarities between *sets of skills* from real-time job ads data. We apply this method to build a job transitions recommender system for identifying transition pathways for workers based on their underlying skill sets. Not only do we accurately predict occupational transitions according to historic job flow data, but we also account for the asymmetric nature of job transitions (the direction of movement between jobs affects its difficulty). Last, we demonstrate the flexibility and performance of our skill set similarity method by building a new technology adoption indicator from skills data, where we showcase the temporal adoption of AI skills in Australian industries. This work relates to the job transition and AI adoption themes of this thesis. As of June 2021, this paper is under publication review with *PLOS ONE*.

Abstract

Job security can never be taken for granted, especially in times of rapid, widespread and unexpected social and economic change. These changes can force workers to transition to new jobs. This may be because new technologies emerge or production is moved abroad. Perhaps it is a global crisis, such as COVID-19, which shutters industries and displaces labor *en masse*. Regardless of the impetus, people are faced with the challenge of moving between jobs to find new work. Successful transitions typically occur when workers leverage their existing skills in the new occupation. Here, we propose a novel method to measure the similarity between occupations using their underlying skills. We then build a recommender system for identifying optimal transition pathways between occupations using job advertisements (ads) data and a longitudinal household survey. Our results show that not only can we accurately predict occupational transitions (Accuracy = 76%), but we account for the asymmetric difficulties of moving between jobs (it is easier to move in one direction than the other). We also build an early warning indicator for new technology adoption (showcasing Artificial Intelligence), a major driver of rising job transitions. By using real-time data, our systems can respond to labor demand shifts as they occur (such as those caused by COVID-19). They can be leveraged by policy-makers, educators, and job seekers who are forced to confront the often distressing challenges of finding new jobs.

6.1 Introduction

In March 2020, COVID-19 caused entire industries to shutter as governments scrambled to 'flatten the curve'. Jobs were lost or subject to an indefinite hiatus; firms went into 'hibernation' to wait out the depressed demand; and governments exercised wartime measures of labor redeployment and wage subsidies of unprecedented scale. All in a matter of weeks.

Labor market shocks, such as those caused by COVID-19, force workers to abruptly transition between jobs. Crises, however, are not the only cause of large-scale job transitions. Structural shifts in labor demand are another major obstacle [7], but usually unfold more gradually. Indeed, technological advances were expected to cause the next wave of major labor market disruptions [87, 299]. The 'future of work' was to be defined by technologies like Artificial Intelligence (AI); technologies that would automate and augment workers, but at the same time transform the requirements of jobs and the demand for labor *en masse* [8, 148, 151].

Despite the impetus, many workers need to transition between jobs. In some countries, such as Australia, labor displacement has increased over the past two decades with relatively high levels of job transitions [308]; a situation exacerbated by COVID-19 [34]. While job turnover is not innately negative and can be a signal of labor market dynamism, it does depend on how efficiently workers transition back into the workforce. Transitioning from one job to another can be difficult or unfeasible when the skills gap is too large [244]. Successful transitions typically involve workers leveraging their existing skills and acquiring new skills to meet the demands of the target occupation [158, 280]. Therefore, transitioning workers successfully at scale requires maximizing the similarity between workers' current skills and their target jobs. Skills, knowledge areas, and capabilities enable workers to achieve tasks required by jobs [242]. We refer to these aspects of human capital as 'skills' throughout this research, and we characterize labor market entities (individual jobs, standardized occupations, industries, etc.) as *sets of skills*.

Here, we propose a novel method to measure the distance between sets of skills from more than 8 Million real-time job advertisements (ads) in Australia from 2012-2020. We call this data-driven methodology SKILLS SPACE. The SKILLS SPACE method enables us to measure the distance between any defined skill sets based on distances at the individual skill level. When two skill sets are highly similar (for example, two occupations), the skills gap is narrow, and the barriers to transitioning from one to the other are low. Drawing from previous work [57, 158, 244, 280], we construct a unique *Job Transitions Recommender System* that incorporates the skill set distance measures together with other labor market data from job ads and employment statistics. This allows us to account for a wealth of labor market variables from multiple sources. The outputs of the recommender system accurately predict transitions between occupations (Accuracy = 76%) and are validated against a dataset of occupational transitions from a longitudinal household survey [130]. While previous studies have analyzed job transitions using the same or similar job ads data [23, 208, 341, 342], they have not accounted for the asymmetries between jobs (please refer to the *Supplementary Information* for a detailed review of the related literature). Our system accounts for the asymmetries between occupations (it is easier to move in one direction than the other), leverages real-time job ads data at the granular skills-level, and accurately recommends occupations and skills that can assist workers looking to transition between jobs based on their personalized skills set. We further demonstrate the flexibility of the SKILLS SPACE method by constructing a leading indicator of Artificial Intelligence (AI) adoption within Australian industries. In our applications of SKILLS SPACE, we are able to both *recommend* transition pathways to workers based on their personalized skill sets and *detect* emerging AI disruption that could accelerate job transitions.

6.2 Materials and methods

6.2.1 Datasets and ground-truth

Job ads data. This research draws on 8,002,780 online job ads in Australia from 2012-01-01 to 2020-04-30, courtesy of Burning Glass Technologies (BGT). This dataset provides unique insights into the evolving labor demands of Australia. It also covers the early periods of the COVID-19 crisis when Australian governments closed 'non-essential' services [4]. To construct this dataset, BGT has systematically collected job ads via web-scraping. This process removes duplicates of job ads posted across multiple job boards or job ads re-posted in short time-frames. They also parse the unstructured job description text through their proprietary systems that extract key features from the advertised job. These features include location, employer, salary, education requirements, experience demands, occupational class, industry classifications, among others. Importantly for this research, the skill requirements have also been extracted (>11,000 unique skills). Here, BGT adopt a broad description of 'skills' to include skills, knowledge, abilities, and tools & technologies. This is slightly different to the more commonly used skills data from O*NET, which defines skills as a series of developed capacities that are categorized into different competencies [326]. There are two major advantages of using BGT job ads data over O*NET skills data: (1) more granular 'skills' data and (2) longitudinal (when used historically) and near-real-time skills data in specific locations. The latter point is particularly important when building a real-time job transitions recommender system to navigate labor crises as they unfold.

Employment statistics. The employment data used for this research is drawn from the 'Quarterly Detailed Labor Force' statistics by the Australian Bureau of Statistics (ABS) [31]. These data represent labor supply features for the 4-digit occupations in the *Job Transitions Recommender System* and include measures of employment levels and hours worked per occupation.

Occupational transitions ground-truth. The Household, Income and Labour Dynamics in Australia (HILDA) Survey is a nationally representative longitudinal panel study of Australian households that commenced in 2001 [130]. It has three main areas of interest: income, labor, and family dynamics. The HILDA survey is in its 18th consecutive year, with the latest available data available from 2018.

Included within the HILDA are data on occupational history and movements of anonymized respondents. We use this data to identify when respondents have changed jobs from one year to another. The occupations are recorded at the 4-digit level from the Australian and New Zealand Standard Classification of Occupations (ANZSCO). This shows the occupation of the previous year and the current year. We use this longitudinal dataset as the ground truth for validating SKILLS SPACE. As the job ads dataset used for this research begins at 2012, we isolate the observations of occupational transitions from 2012 to 2018 (the latest available year). This results in a sample of 2,999 occupational transitions in Australia.

6.2.2 Measuring skill similarity

To measure the distance between occupations (or other skill groups), we first measure the pairwise distance between individual skills (6,981 skills in 2018) in jobs ads for each calendar year from 2012-2020. Intuitively, two skills are similar when they are simultaneously important for the same set of job ads. We measure the importance of a skill in a job ad using an established measure called 'Revealed Comparative Advantage' (RCA – Eq. (6.1)) that has been applied across a range of disciplines, such as trade economics [185, 330], identifying key industries in nations [307], detecting the labor polarization of workplace skills [14], and adaptively selecting occupations according to their underlying skill demands [120]. RCA normalizes the total share of demand for a given skill across all job ads. We then calculate the pairwise skill similarity between each skill using Eq. (6.2) as implemented by Alabdulkareem et al. [14] and again by Dawson et al. [120]. These individual skill distances form the basis for measuring the distance between sets of skills.

To measure the distance between every skill for each year in the dataset, we start by removing extremely rare skills. Here, we select skills with a posting frequency count >= 5, which represent ~ 75% of all skills (see *Supplementary Information* for more details). Let $\mathbb{S}$ be the set of all skills and $\mathbb{J}$ be the set of all job ads in our dataset. We measure the similarity between two individual skills s_1 and s_2 ($s_1, s_2 \in \mathbb{S}$) using a methodology proposed by Alabdulkareem et al. [14]. First, we use the *Revealed Comparative Advantage*

(RCA) to measure the importance of a skill s for a particular job ad j:

$$RCA(j,s) = \frac{x(j,s) / \sum_{s' \in \mathbb{S}} x(j,s')}{\sum_{j' \in J} x(j',s) / \sum_{j' \in \mathbb{J}, s' \in \mathbb{S}} x(j',s')} \tag{6.1}$$

where $x(j,s) = 1$ when the skill s is required for job j, and $x(j,s) = 0$ otherwise; $RCA(j,s) \in \left[0, \sum_{j' \in \mathbb{J}, s' \in \mathbb{S}} x(j',s')\right], \forall j,s$, and the higher $RCA(j,s)$ the higher is the comparative advantage that s is considered to have for j. Visibly, $RCA(j,s)$ decreases when the skill s is more ubiquitous (i.e. when $\sum_{j' \in \mathbb{J}} x(j',s)$ increases), or when many other skills are required for the job j (i.e. when $\sum_{s' \in \mathbb{S}} x(j,s')$ increases). Next, we measure the similarity between two skills based on the likelihood that they are both effectively used in the same job ads. Formally:

$$\theta(s_1,s_2) = \frac{\sum_{j \in \mathbb{J}} e(j,s_1) \cdot e(j,s_2)}{max\left(\sum_{j \in \mathbb{J}} e(j,s_1), \sum_{j \in \mathbb{J}} e(j,s_2)\right)} \tag{6.2}$$

where $e(j,s)$ is the effective use of a skill in a job, defined as $e(j,s) = 1$ when $RCA(j,s) \geq 1$ and $e(j,s) = 0$ otherwise. Note that $\theta(s_1,s_2) \in [0,1]$, a larger value indicates that s_1 and s_2 are more similar, and it reaches the maximum value when s_1 and s_2 always co-occur (i.e. they never appear separately) while $e(j,s_1) = 1$ and $e(j,s_2) = 1, \forall j \in \mathbb{J}$. Visibly, $\theta(s_1,s_2)$ is based on the co-occurrence of skills when both s_1 and s_2 are simultaneously important for the job ads. Therefore, θ measures when two skills are effectively used together – i.e., it measures similarity as in "complementary", not as in "replaceable".

SKILLS SPACE Method

Next, we use the pairwise skill distances to measure the distance between *sets of skills*, which we refer to as SKILLS SPACE. Here, a set of skills can be arbitrarily defined, such as an occupation, an industry, or a personalized skills set. Intuitively, two sets of skills are similar when their most important skills are similar. We first introduce a measure of skill importance within a skill set as the mean RCA over all the job ads pertaining to the skill set. Assume a job ads grouping criterion exists, for example, job ads pertaining to an occupation, a company, or an industry. We obtain the job ads set $\mathscr{J} \subset \mathbb{J}$ and the set of skills $\mathscr{S} \subset \mathbb{S}$ occurring within $j \in \mathscr{J}$. We denote $\mathscr{J}$ as the set of job ads associated with the skill set $\mathscr{S}$. We measure the importance of skill s for $\mathscr{S}$ (and implicitly for $\mathscr{J}$) as the mean RCA over all the job ads relating to the skill set $\mathscr{S}$. Formally,

$$w(s,\mathscr{S}) = \frac{1}{|\mathscr{J}|}\sum_{j\in\mathscr{J}} RCA(j,s) \tag{6.3}$$

Next, we propose a method to measure the distance between *sets of skills*. For example, suppose there are two jobs that we can define by their underlying skill demands. Both jobs have their unique set of skills, and each individual skill has its own relative importance to the specific job, as calculated by Eq. (6.3). Intuitively, the two jobs are similar when their most important skills (i.e., their 'core' skills) are similar. This is achieved by computing the weighted pairwise skill similarity between the individual skills of each job (using Eq. (6.2)), where the weights correspond to the skill importance (defined by Eq. (6.3)). This returns a single similarity score between the two skill sets corresponding to the two jobs. Formally, let $\mathscr{S}_1$ and $\mathscr{S}_2$ be two sets of skills, and $\mathscr{J}_1$ and $\mathscr{J}_2$ their corresponding sets of job ads. We define Θ the similarity between $\mathscr{S}_1$ and $\mathscr{S}_2$ as the weighted average similarity between the individual skills in each set, where the weights correspond to the skill importance in their respective sets. Formally,

$$\Theta(\mathscr{S}_1,\mathscr{S}_2) = \frac{1}{C}\sum_{s_1\in\mathscr{S}_1}\sum_{s_2\in\mathscr{S}_2} w(s_1,\mathscr{S}_1)w(s_2,\mathscr{S}_2)\theta(s_1,s_2) \tag{6.4}$$

where $C = \sum_{s_1\in\mathscr{S}_1}\sum_{s_2\in\mathscr{S}_2} w(s_1,\mathscr{S}_1)w(s_2,\mathscr{S}_2)$. Similar to θ defined in Eq. (6.2), Θ is a similarity measure (higher means more similar) and $\Theta(\mathscr{S}_1,\mathscr{S}_2)\in[0,1]$. Note that Θ is a compound measure based on θ, which in turn measures the complementarity of two skills (see prior discussion and interpretation of θ). As a result, $\Theta(\mathscr{S}_1,\mathscr{S}_2)$ measures the complementarity of two skill sets. The interpretation we use in the rest of this paper is that "when Θ is high, an individual with $\mathscr{S}_1$ can more readily fulfil the skill requirements of $\mathscr{S}_2$". We use Θ as a key feature in our job transitions recommender system. The setup and details of this system now follow.

6.2.3 Job Transitions Recommender System setup

We construct the job transitions recommender system as a binary machine learning classifier using XGBoost – an implementation of gradient boosted tree algorithms, which has achieved state-of-the-art results on many standard classification benchmarks with medium-sized datasets [1]. The XGBoost algorithm is a linear combination of decision trees where each subsequent tree attempts to reduce the errors from its predecessor. This allows for the next tree in the series 'learn' from the errors of the previous tree with the goal of making more accurate predictions. In our application of the XGBoost algorithm, the system 'learns' from the input labor market data, which are independent variables (or features). It is then 'trained' against historic examples of occupational transitions that did occur (positive examples) and did not occur (negative examples), which are the dependent variables (or ground-truth). As is standard in machine learning practice, we reserve a 'test set' of observations for evaluation, where we apply the trained model to make predictions about whether a transition occurred or not (hence, binary) by only observing the features. This setup allows us to predict the probability of an occupational transition from the 'source' to the 'target' occurring (positive example) or not (negative). Here, we use the job-to-job transitions data from the HILDA dataset [130] (described above). We then randomly simulate an alternate sample of transitions where we maintain the same 'source' occupations and randomly select 'target' occupations (called 'Random Sample'). This produces a balanced dataset of 5,998 positive and negative occupational transitions. We then associate each 'source' and 'target' observation with their temporal pairwise distance measure using the SKILLS SPACE method. However, the SKILLS SPACE measures are symmetric, and job transitions are known to be asymmetric [57, 244, 289]. Therefore, to represent the asymmetries between job transitions, we add a range of explanatory features to each 'source' and 'target' occupation. These occupational features include their SKILLS SPACE pairwise distance measures and other variables, such as years of education required, years of experience demanded, and salary levels, from employment statistics ('Labor Supply') and job ads data ('Labor Demand' – see *Supplementary Information* for full list of features).

Like most machine learning algorithms, XGBoost has a set of hyper-parameters – parameters related to the internal design of the algorithm that cannot be fit from the training data. The hyper-parameters are usually tuned through search and cross-validation. In this work, we employ a Randomized-Search [65] which randomly selects a (small) number of hyper-parameter configurations and performs evaluation on the training set via cross-validation. We tune the hyper-parameters for each learning fold using 2500 random combinations, evaluated using a 5 cross-validation. We train the models on 80% of the observations, leaving aside 20% of the data for testing, which we randomly seed. We repeat the process 10 times for each feature model configuration and change the random seed to select a new testing sample, which provides us with the standard deviation bars seen in Fig. 6.3.

6.2.4 Constructing a leading indicator of AI adoption

Adapting the SKILLS SPACE method, we develop a leading indicator for emerging technology adoption and potential labor market disruptions based on skills data, using AI as an example. We select AI because of its potential impacts on transforming labor tasks and accelerating job transitions [87, 148, 151]. Our indicator temporally measures the similarity between a dynamic set of a AI skills against the 19 Australian industry skill sets from 2013-2019.

To create these yearly sets of top AI skills, we first select a sample of core 'seed skills' that are highly likely to remain important to AI over time – here we selected 'Artificial Intelligence', 'Machine Learning', 'Data Science', 'Data Mining', and 'Big Data' as the seed skills. This set of seed skills represents $\mathscr{S}$ from Eq. (6.3) as opposed to a grouping criterion, such as an occupation or industry. In this case, $\mathscr{J}$ is not defined, and we measure the importance of a skill as the average θ similarity to the seed skills. Repeating this process temporally allows us to build dynamic skill sets. We then use Eq. (6.2) to measure the similarity (θ) of each seed skill to every other skill in a given year. By calculating the average θ for every skill relative to the seed skills, we return an ordered list of skills with the highest similarity. This process is repeated for each calendar year from 2013-2019 where we select the top 100 skills for each year. The skill similarity approach allows us to build an adaptive list of AI skills that captures evolving skill demands. This is especially important for a skill area like AI, where the skill demands are changing very quickly. For example, 'TensorFlow' (a Deep Learning framework) emerged as a skill in November 2015 and has since become among the fastest-growing AI skills. The AI skill lists we create can detect the importance of 'TensorFlow' in 2016, whereas a static list pre-defined before 2016 would have missed these important changes to AI skill demands.

Having constructed temporal sets of AI skills, we then measure the yearly similarities between the AI skillsets and the skill sets of Australia's 19 major industries – classified according to the Australian & New Zealand Standard Industrial Classification (ANZSIC) Division level. Using the SKILLS SPACE method, we construct each industry as a set of skills for every year and use Eq. (6.4) to calculate similarity to the yearly AI skill sets. This allows us to observe and compare the extent to which AI skills have diffused throughout industries and the relative importance of AI skills to these industries. The advantages of using this skill similarity approach as opposed to *ad hoc* skill counts from pre-defined skills are twofold. First, we create dynamic sets of skills that capture evolving skill demands. Second, we account for skill importance within individual job ads by normalizing for high-occurring skills (see *Supplementary Information* for more details).

6.3 Results & Discussion

6.3.1 Skill similarity results

Fig. 6.1A shows the two-dimensional skill distance embeddings for the top 500 skills by posting frequency in 2018. Here, each marker represents an individual skill that is colored according to one of 13 clusters of highly similar skills, using the K-Medoids clustering algorithm. By using a triplets method for dimensionality reduction [16], we are able to preserve the global structure of the embedding (global structure = 98%). That is, two markers are depicted closer together when their corresponding skills are more similar (i.e., have low distance). This provides useful insights, highlighting that specialized skills (such as 'Software Development' and 'Healthcare') tend to lay toward the edges of the embedding, whereas more general and transferable skills lay toward the middle, acting as a 'bridge' to specialist skills. Highly similar skills cluster closely together; for example, the 'Software Development' cluster (see inset) regroups programming skills such as scripting languages 'Python', 'C++', and 'HTML5'. It is important to measure the similarity between jobs based on their underlying skills because workers leverage their existing skills to make career changes [88].

6.3.2 SKILLS SPACE results

In Fig. 6.1B, the markers depict groups of skills that correspond to individual occupations, using the official Australian standard (at the 6 digit level – see *Supplementary Information* for more details). To visualize the distance between occupations, we use the same dimensionality reduction technique as for individual skills in Fig. 6.1A. Each occupation is colored on a scale according to their automation susceptibility, as calculated by Frey and Osborne [151] – dark blue represents low-risk probability, and dark red shows high-risk probability over the coming two decades. As seen in Fig. 6.1B and the magnified inset, similar occupations lie close together on the map. Furthermore, occupations at low risk of automation tend to be characterized by non-routine, interpersonal, and/or high cognitive labor tasks [41]. In contrast, occupations at high risk of automation require routine, manual, and/or low cognitive labor tasks. For example, in the inset of Fig. 6.1B, a 'Sheetmetal Trades Worker' is deemed to be at high risk of labor automation

Figure 6.1: **Measuring the distance between skills and occupations.** (A) Shows a two-dimensional embedding of the 2018 skill distances, where the top 500 skills by posting frequency are visualized. Each marker represents an individual skill colored according to 13 clusters using K-Medoids clustering. As observed in the 'Software Development' inset, highly similar skills cluster together. Additionally, the specialized skill clusters, such as 'Software Development' and 'Healthcare' tend to lay toward the edges; whereas the more general and transferable skills lay toward the middle of the embedding and act as a 'bridge' to specialist skills. These individual skill distances form the basis of measuring the distance between sets of skills. (B) We leverage SKILLS SPACE to measure the distance between official Australian occupations at the 6-digit level (characterized by their skill sets) in 2018. The markers represent individual occupations, colored by technological labor automation risk, as calculated by Frey & Osborne. Occupations that require lower levels of routine, manual and/or low cognitive labor tasks tend to be at higher risk (colored darker red); whereas occupations characterised by non-routine, interpersonal, and/or high cognitive labor tasks are at lower risk (colored darker blue) over the next two decades.

(82% probability) due to high levels of routine and manual labor tasks required by the occupation. However, a 'Sheetmetal Trades Worker' skillset demands are highly similar to an 'Industrial Designer', which is considered at low risk of labor automation over the coming two decades (4% probability). Therefore, an 'Industrial Designer' represents a transition opportunity for a 'Sheetmetal Trades Worker' that leverages existing skills and protects against potential risks of technological labor automation.

6.3.3 Validation of SKILLS SPACE distance

We validate the link between the SKILLS SPACE and job transitions by conducting paired t-tests, as illustrated in Fig. 6.2. Here, we use a longitudinal household survey dataset containing actual job transitions [130] (called the 'True Sample'). Each occupational pair ('source' to 'target') is labeled with its SKILLS SPACE distance for the given year. We randomly simulate an alternate transition sample by maintaining the same 'source' occupations and randomly selecting 'target' occupations (called 'Random Sample'). Our objective is to determine whether the differences in SKILLS SPACE distance between the 'True Sample' and the 'Random Sample' are statistically significant. First, we test the differences of the two samples, including *all* occupational transitions. We find that the differences between the two samples are statistically significant (t-statistic = 16.272, p-value = 2.707×10^{-58}, Cohen's D effect size = 0.42) (see *Supplementary Information*). However, one-third of the 'True Sample' transitions are to another job but are classified as the same occupation. Therefore, we perform a second test only on transitions between different occupations, i.e., we remove all observations where the 'source' and 'target' are identical. Again, the differences between the 'True Sample' and the 'Random Sample' are statistically significant (t-statistic = 4.514, p-value = 6.535×10^{-6}, Cohen's D effect size = 0.14), as illustrated in Fig. 6.2A. We repeat the procedure 100 times: we generate 100 new 'Random Samples', and we perform the statistical test for each of them. 87% of these tests are statistically significant, as shown in Fig. 6.2B. These results provide evidence that the SKILLS SPACE distance measure is representative of actual job transitions.

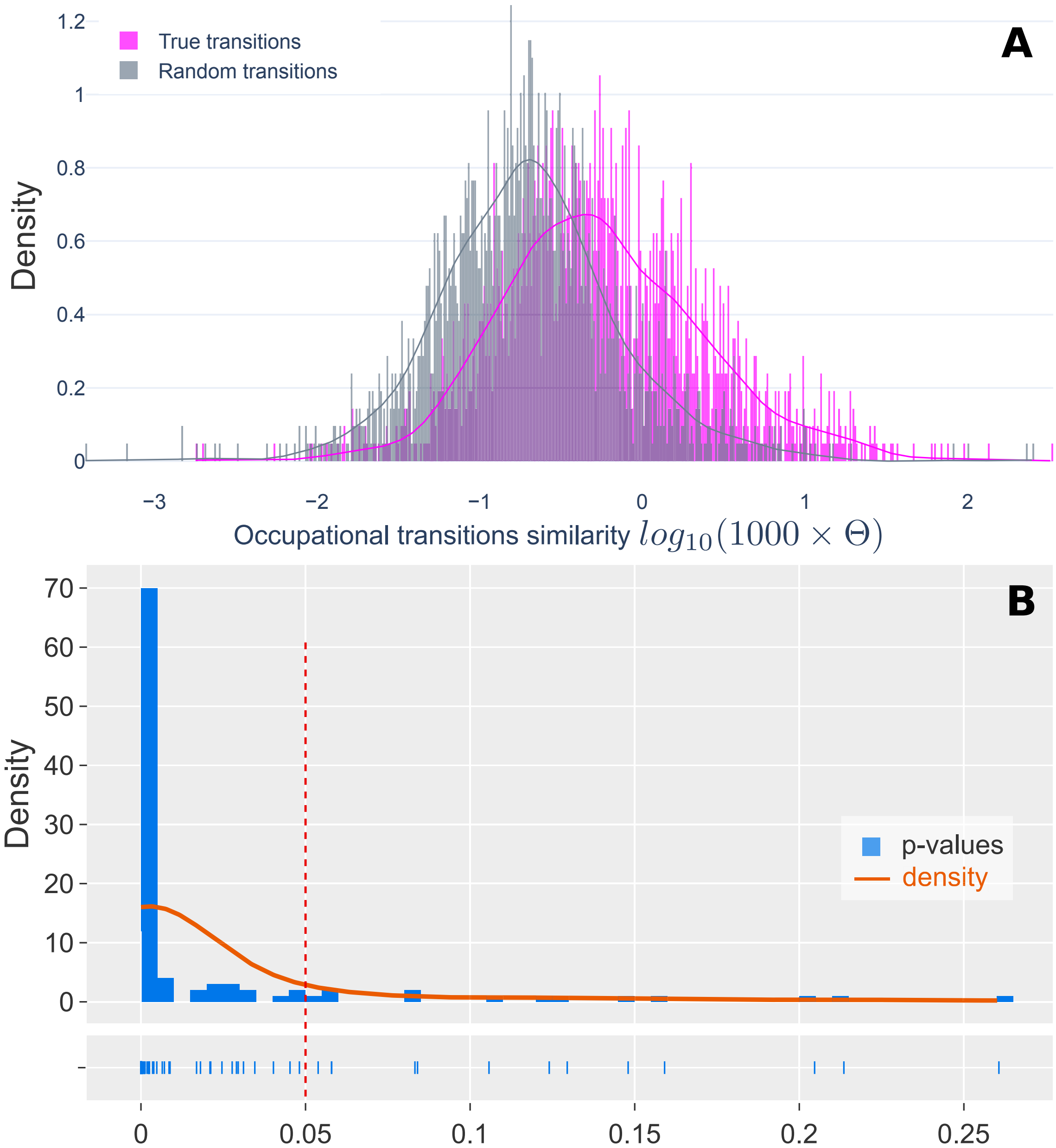


Figure 6.2: **The SKILLS SPACE is statistically significant in representing job transitions. (A)** The x-axis shows the log-transformed SKILLS SPACE distance for a 'True' sample of actual transitions (shown in magenta color) and a 'Random' sample of simulated transitions (shown in gray color). Each Random transition is paired with an Actual transition: it shares the same 'source' occupation as the Actual transition but the 'target' occupation is randomly selected and is different to the Actual observation. The difference between the two populations is statistically significant (paired t-test, t-statistic = 4.514, p-value = 6.535×10^{-6}, Cohen's D effect size = 0.14). **(B)** We repeat the procedure 100 times: we generate 100 'Random' populations, and we perform the statistical testing for each. The figure shows the histogram (density and rug plot) of the 100 obtained p-values, 87 of which are lower than 0.05.

6.3.4 Job Transitions Recommender System

Job transitions, however, are *asymmetric* [57, 244, 289] – the direction of the transition affects the difficulty. Therefore, transitions are determined by more than the symmetric distance between skill sets; other factors, such as educational requirements and experience demands, contribute to these asymmetries. We account for the asymmetries between job transitions by constructing a machine learning classifier framework that combines the SKILLS SPACE distance measures with other labor market features from job ads data and employment statistics (discussed in *Job Transitions Recommender System setup*). We then apply the trained model to predict the probability for every possible occupational transition in the dataset – described as the transition probability between a 'source' and a 'target' occupation. This creates the *Transitions Map*, for which a subset of 20 occupations can be seen in Fig. 6.3. The colored heatmap shows the transition probabilities ('source' occupations are in columns and 'targets' are in rows). Dark blue represents higher transition probabilities, and lighter blue shows lower probabilities, where the asymmetries between occupation pairs are clearly observed. For example, a 'Finance Manager' has a higher probability of transitioning to become an 'Accounting Clerk' than the reverse direction. Moreover, transitioning to some occupations is generally easier (for example, 'Bar Attendants and Baristas') than others ('Life Scientists'). The dendrogram illustrates the hierarchical clusters of occupations where there is a clear divide in Fig. 6.3 between service-oriented professions and manual labor occupations.

For validation, we train various classifier models with different feature configurations and identify three main findings. *First*, as seen in Fig. 6.4, the models that incorporate the distance measures with all of the other occupational features ('All Features') consistently achieve the highest accuracy for occupational transitions (Accuracy = 76%). This feature setup achieves higher results than models that only use the 'Labor Demand + Labor Supply' features (Accuracy = 74%) or the distance measures alone (Accuracy = 73%). To further validate these findings, we conduct an ablation test where each feature is iteratively removed from the feature set and the model is re-trained – the model configurations with lower performance indicate higher feature importance. Here, the exclusion of the SKILLS SPACE distance measure caused the greatest decline in performance, therefore reiterating its predictive power. We also conduct a feature importance analysis, which again shows that the SKILLS SPACE distance measure is the most important feature for predicting transitions (see *Supplementary Information* for further details). *Second*, the standard deviation of accuracy over repeated trials decreases when all features are combined (as seen with the spread of the performance bars in Fig. 6.4). This shows that

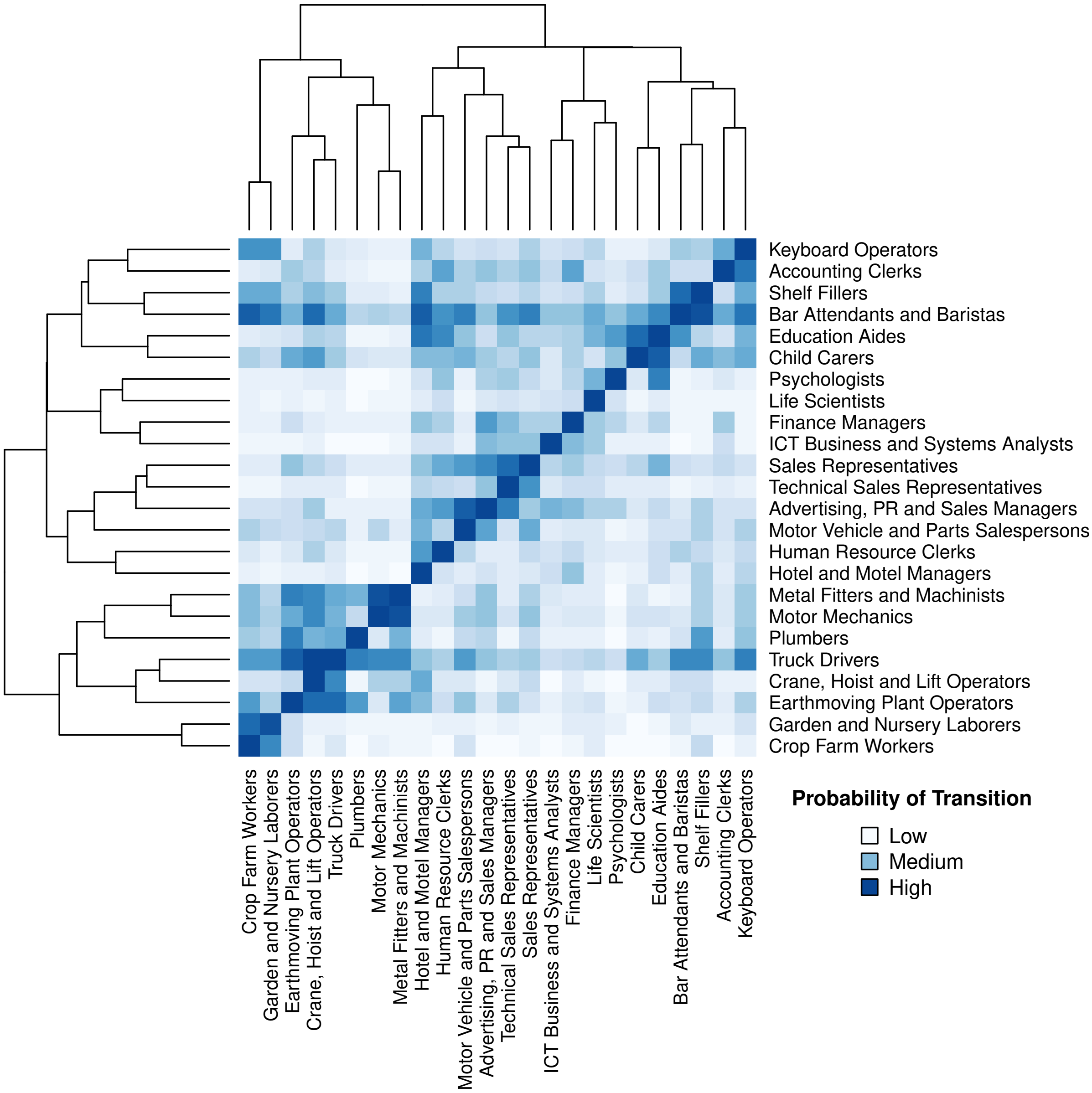


Figure 6.3: Validation and the *Transitions Map*. Visualizes a subset of the *Transitions Map*, where 20 occupations and their pairwise transition probabilities can be observed. In this visualization, transitions occur from columns to rows, and dark blue depicts high transition probabilities, and white depicts low probabilities. While job transitions to the same occupation yield the highest probabilities (dark blue diagonal squares), it is clear that transitions are asymmetric. The dendrogram highlights how similar occupations cluster together, where there is a clear divide between services and manual labor occupations.

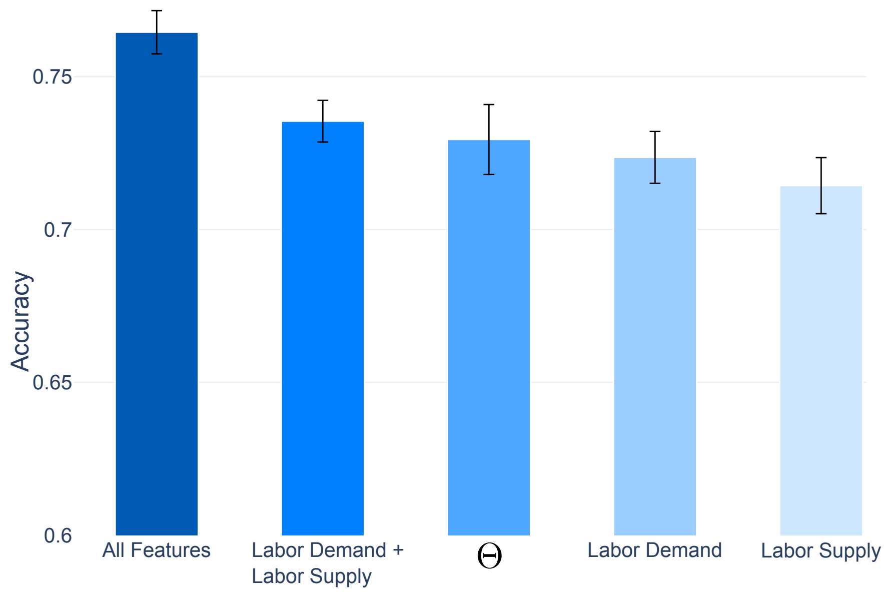


Figure 6.4: The prediction accuracy scores of the different classifier model feature configurations. The highest performance is achieved when 'All Features' are incorporated in the classifier models to predict occupational transitions (76%). Moreover, by incorporating all features, the standard deviation decreases (shown by the performance bars), which highlights the complementarity of the combined features and the ability to now account for the asymmetry between jobs.

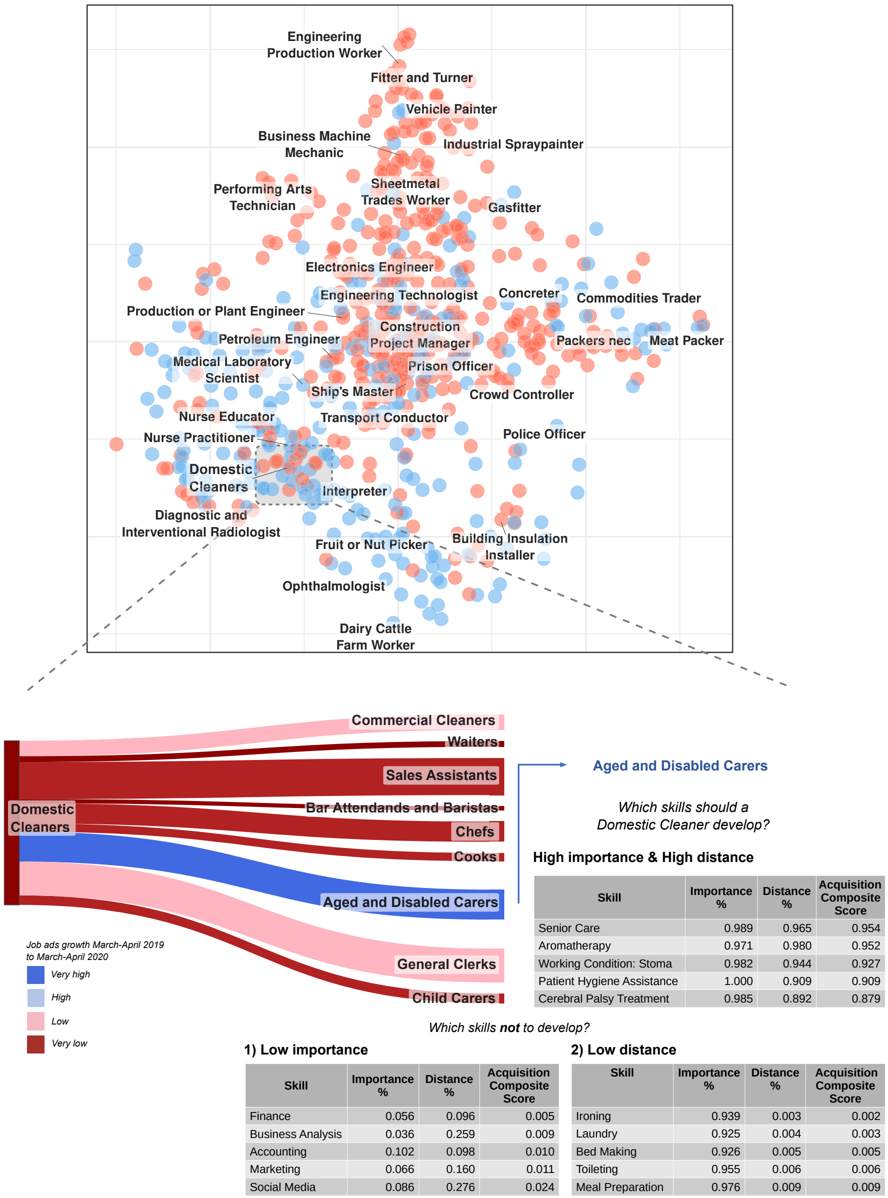
the SKILLS SPACE distance measures and the occupational features are complementary in predicting job transitions. *Third*, and most important, is that by combining all features, we can construct the asymmetries between occupations. While transitioning to a job in the same occupation yields the highest probabilities (as seen by the dark blue diagonal line in Fig. 6.3), the occupational features add asymmetries between occupational pairs, such as accounting for asymmetries in education and experience requirements.

6.3.4.1 Recommending Jobs and Skills.

The *Transitions Map* provides the basis for making qualified job transition recommendations. We call this the *Job Transitions Recommender System*. In Fig. 6.5, we showcase its usage in the context of a labor market crisis (i.e. COVID-19). We highlight the example of 'Domestic Cleaners', an occupation that experienced significant declines in labor demand and employment levels during the crisis (see *Supplementary Information*).

During the 'first wave' of COVID-19 in March 2020, the Australian Government enforced social-contact and mobility restrictions on 'non-essential services' to slow the spread of the virus [174]. Many occupations within these 'non-essential' services were unable to trade and perform their duties, forcing some workers to try and transition to another job. In the upper panel of Fig. 6.5, we visualize the 'essential' and 'non-essential' occupations on the *Transitions Map*. The placement of the occupations is identical to Fig. 6.1B and we highlight the 'essential' occupations as the blue nodes and the 'non-essential' occupations are the red nodes using the classifications developed by Faethm AI [139]. We observe that the cluster of medical occupations (bottom-left of the map) are deemed as 'essential', as are most of the food production workers (bottom). Here, we select 'Domestic Cleaners' as an example of a 'non-essential' occupation and use the *Transitions Map* to recommend the occupations with the highest transition probabilities in the bottom panel of Fig. 6.5. These are the nodes on the right side of the flow diagram in Fig. 6.5, ordered in descending order of transition probability. Segment widths show the labor demand for each of the recommended occupations during the COVID-19 period (measured by posting frequency). The segment colors represent the percentage change of posting frequency during March and April 2019 compared to the same months in 2020; dark red indicates a big decrease in job ad posts, and dark blue indicates a big increase. The first six transition recommendations for 'Domestic Cleaners' have all experienced negative demand, which is unsurprising given that they were also deemed as 'non-essential' services. However, the seventh recommendation, 'Aged and Disabled Carers', has significantly grown in demand during the COVID-19 period, and there is a high number of jobs advertised. 'Aged and Disabled Carers' is both an 'essential' and a high-demand occupation; therefore, it makes sense to select 'Domestic Cleaner' as the target occupation for transitioning into.

We take the *Job Transitions Recommender System* a step further by incorporating skill recommendations. Transitioning to a new occupation generally requires developing new skills under time and resource constraints. Therefore, workers must prioritize which skills to develop. We argue that a worker should invest in developing a skill when (1) the skill is important to the target occupation *and* (2) the distance to acquire the skill is large (that is, it is relatively difficult to acquire). Formally, for a target skill (i.e., a skill in the 'target' occupation), we compute its *importance* to the target occupation and its *distance* to the source occupation. We calculate *skill importance* as the mean RCA for the skill across all job ads within the target occupation using Eq. (6.3). We calculate *skill distance* as the distance between the target skill and 'source' occupation skill set as $1-\Theta(S_1, S_2)$ using Eq. (6.4) (i.e., the 'target' skillset (S_2) is made out of a single skill: the target skill). Finally, we build the *acquisition composite score* as the multiplication of importance and distance, transformed as percentiles to account for variation in scale.



High importance & High distance

Skill	Importance %	Distance %	Acquisition Composite Score
Senior Care	0.989	0.965	0.954
Aromatherapy	0.971	0.980	0.952
Working Condition: Stoma	0.982	0.944	0.927
Patient Hygiene Assistance	1.000	0.909	0.909
Cerebral Palsy Treatment	0.985	0.892	0.879

1) Low importance

Skill	Importance %	Distance %	Acquisition Composite Score
Finance	0.056	0.096	0.005
Business Analysis	0.036	0.259	0.009
Accounting	0.102	0.098	0.010
Marketing	0.066	0.160	0.011
Social Media	0.086	0.276	0.024

2) Low distance

Skill	Importance %	Distance %	Acquisition Composite Score
Ironing	0.939	0.003	0.002
Laundry	0.925	0.004	0.003
Bed Making	0.926	0.005	0.005
Toileting	0.955	0.006	0.006
Meal Preparation	0.976	0.009	0.009

Figure 6.5: Here, we demonstrate the *Job Transitions Recommender System* using the 'Domestic Cleaner' occupation as an example – an occupation classified as a 'non-essential' occupation and that has experienced significant declines during the beginning of the COVID-19 outbreak in Australia. (upper panel) The two-dimensional space of occupations (see Fig. 6.1) with 'essential' occupations in blue markers and 'non-essential' occupations in red markers. (lower panel) We first use the *Transitions Map* to calculate the occupations with the highest transition probabilities (other than itself). These are the nodes on the right side of the flow diagram in the bottom panel of the figure, where the link colors depict posting frequency percentage change from March-April 2019 to March-April 2020. The link widths represent the posting frequency volume of March-April 2020 to indicate labor demand. The top six occupations have all experienced significant declines during the COVID-19 period; however, the seventh recommendation, 'Aged and Disabled Carers', is experiencing significant labor demand growth. 'Aged and Disabled Carers' were also classified as an 'essential' occupation during COVID-19 in Australia. We select this as the target occupation and then make personalized skill recommendations. We argue that workers trying to transition to another occupation should invest time and resources into skill development when (1) the skills are of **high importance** *and* (2) there is a **high distance** to acquire the skill. Conversely, workers should *not* focus on skill development if (1) the skills are **low importance** *or* (2) there is a **low distance** to acquire the skill.

In the case of the 'Domestic Cleaner' in Fig. 6.5 (lower panel), the skills with the highest acquisition composite score for the transition to 'Aged and Disabled Carer' are specialized patient care skills, such as 'Patient Hygiene Assistance'. Conversely, the reasons not to develop a skill are when (1) the skill is not important *or* (2) the distance is small to the target occupation. Fig. 6.5 (lower panel) shows that while some 'Aged and Disabled Carer' jobs require 'Business Analysis' and 'Finance' skills, these skills are of low importance for the 'Aged and Disabled Carer' occupation, so they should not be prioritized. Similarly, skills such as 'Ironing' and 'Laundry' are required by 'Aged and Disabled Carer' jobs, but the distance is small, so it is likely that either a 'Domestic Cleaner' already possesses these skills or they can easily acquire them. Visibly, for both of the latter cases, the acquisition composite score takes low values.

6.3.5 A leading indicator of AI adoption

Emerging technologies can change the demand for labor and accelerate forced job transitions by disrupting labor markets [79]. However, in order for emerging technologies to have these effects, they must first be widely adopted by firms across many industries. In this sense, technology adoption rates are a precursor to the societal impacts that they impose, such as widespread changes to labor demand and accelerated job transitions. Measuring technology adoption, however, can be difficult as it often depends on the private activities of firms and is influenced by a range of factors (see *Supplementary Information*). Therefore, measuring the drivers that enable emerging technology adoption (see *Supplementary Information*) can provide leading indicators of adoption rates. One major driver is the availability of skilled labor. Firms that can readily access workers with relevant skills are able to make productive use of the emerging technologies and accelerate their adoption rates, particularly in the early stages of technology growth [68]. SKILLS SPACE offers a useful method for identifying the extent of specific skill gaps of firms within industries. As an industry's skills set becomes more similar to the skills of an emerging technology, the skills gap is narrowed, and the barriers to adopting the emerging technology are reduced. When access to relevant skilled labor is plentiful, the labor requirements enabling technological adoption can be readily achieved and help accelerate adoption rates. Therefore, measuring temporal levels of skill set similarity for an emerging technology within an industry provides a useful leading indicator of technology adoption over time. These measures offer early detection signals of emerging technology adoption and the changing skill requirements that could cause labor disruptions within industries, such as forced job transitions.

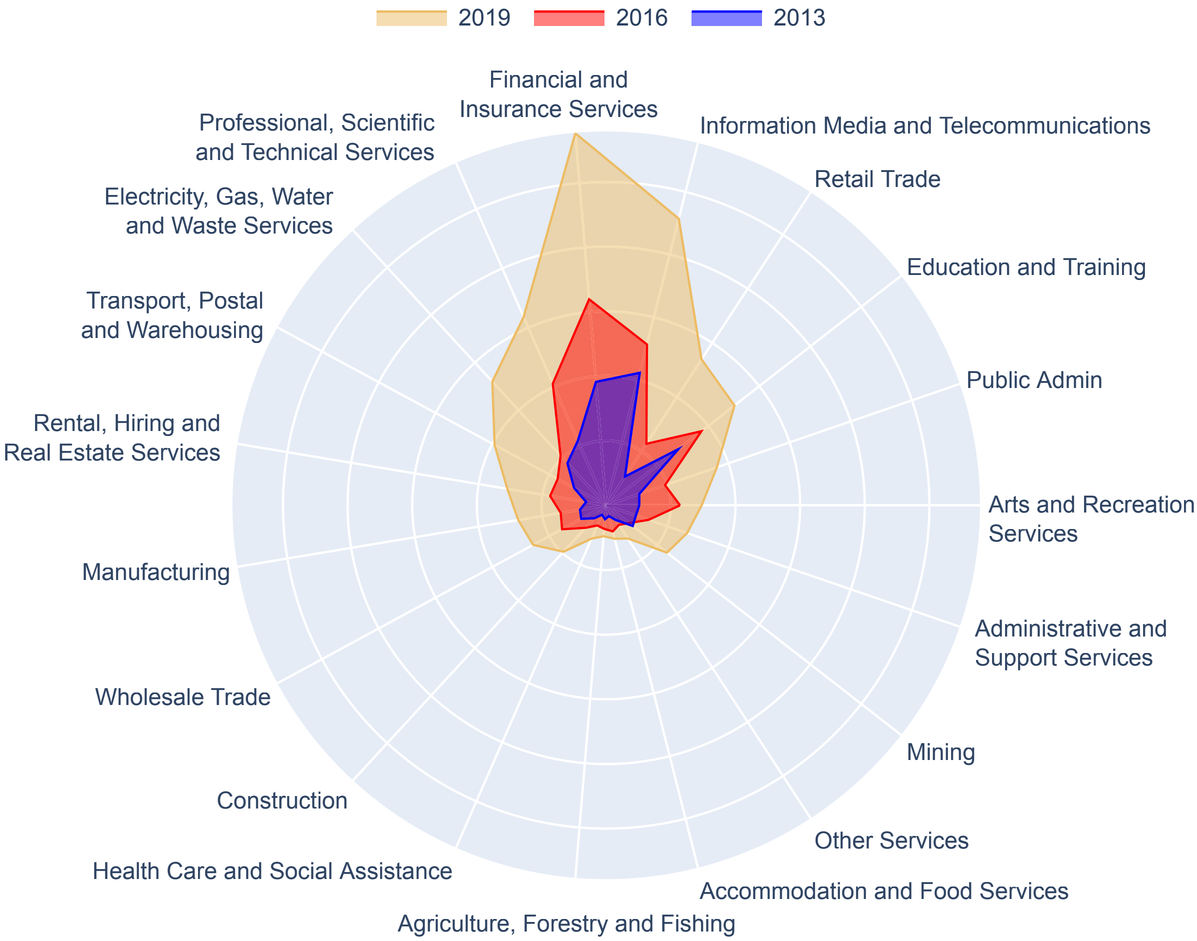


Figure 6.6: By applying SKILLS SPACE, we measure the yearly similarities between adaptive sets of AI skills against each of the 19 Australian industries from 2013-2019. As industry skill sets become more similar to AI skills, the colored area of the radar chart expands. All industries have increased their similarity levels to AI skills, albeit at different rates. We argue that higher levels of AI similarity indicate AI skills are becoming more important to firms within an industry and that the skills gap to acquiring AI skills is narrowed. Access to these skills accelerates the rate of firms adopting AI and making productive use of the technologies, which offers a leading indicator of AI adoption and potential labor disruptions within these industries.

Fig. 6.6 shows that all Australian industry skill sets have grown in similarity to AI skills from 2013 to 2019 – illustrated by the expanding colored areas. This highlights the growing importance of AI skills across the Australian labor market. However, the rates of similarity are unequally distributed. Some industries – such as 'Finance and Insurance Services' and 'Information Media and Telecommunications' – command much higher rates of AI skill similarity. This indicates that not only are firms within these industries increasingly demanding AI skills but also that the AI skills gaps within these industries are much smaller. Also noteworthy are the differences in growth rates toward AI skill similarity. As clearly seen in Fig. 6.6, AI skill similarity has rapidly grown for some industries and more modestly for others. For instance, 'Retail Trade' has experienced the highest levels of growth in similarity to AI skills, increasing by 407% from 2013 to 2019. The majority of this growth has occurred recently, which coincides with the launch of Amazon Australia in 2017 [204]. Since then, Amazon has swiftly hired thousands in Australia.

By adapting the SKILLS SPACE method, we develop a leading indicator that detects AI adoption from real-time job ads data. Such a measure can act as an 'early warning' signal of forthcoming labor market disruptions and accelerated job transitions caused by the growth of AI. This indicator can assist policy-makers and businesses to robustly monitor the growth of AI skills (or other emerging technologies), which acts as a proxy for AI adoption within industries (or other labor market groups).

6.3.6 Limitations

We acknowledge several limitations of the SKILLS SPACE method and the results presented in this paper. First, there are *data limitations* from both the household survey data and the job ads data. The job transitions drawn from the HILDA panel dataset are a relatively small sample, with 2,999 job transitions from 2012-2018. While these observations come from the high-quality HILDA dataset, which is Australia's pre-eminent and representative household survey [130], it is nonetheless a relatively small sample to train a machine learning system against. A small sample enlarges the risks of biases emerging as the recommender system is dependent on a relatively small sample of observations to 'learn' from and make predictions about future job transitions. Longitudinal household surveys also suffer from panel attrition, including HILDA [336]. However, this is only a minor issue for this study, as yearly job transitions are treated as independent observations. Our methods are not dependent on the longitudinal career pathways of individuals. With regards to the job ads data, we only had access to the Australian job ads dataset. As a result, our analyses and results are specific to the Australian labor market. However, this is a feature of our work rather than a limitation, as it allows to contextualize the analysis to geographical and temporal labor markets – it has been shown that labor markets can be highly contextual [239]. One can easily leverage our methods to produce results for other countries by applying equivalent country-specific labor market data from job ads, employment statistics, and occupational transitions.

Second, the results presented in this paper have been *aggregated to the occupational level*. That is, the explanatory features have been grouped by their 4-digit occupations, such as median salary and average education for a given occupation (see the *Supplementary Information* for a full list of the features). Consequently, the job transition predictions in this paper are made at the grouped 4-digit occupational level for demonstration purposes, which does not differentiate within the same occupations. However, the flexibility of the methods presented in this paper can be applied at the individual level (or another arbitrary grouping), given the availability of appropriate data sources.

Third, there can be many factors that cause individuals to transition between jobs [57], beyond those used as explanatory variables in this study. For example, it has been shown that personality profiles are predictive of different occupational classes [202]. Therefore, it is plausible that personality traits and values could influence not only the willingness to move between jobs but also the types of job transitions. Similarly, there exists a Markovian assumption that a worker is described by the set of skills in their current job, therefore ignoring their past work experience and education. Other factors such as competitive dynamics within specific industries and labor markets, macroeconomic conditions, and changes to individuals' household finances can all influence people transitioning between jobs. Future work could look to incorporate these additional variables to help further explain and predict job transitions.

Last, we must acknowledge the risks of biases emerging from applying mechanical algorithms to 'learn' from historical examples and make consequential recommendations to people, such as suggested career pathways. If the data used to train a machine learning system contains biases, then the predictions generated by the system are likely to reflect these biases. For example, there are structural biases in labor markets that influence employment outcomes, including biases based on gender [83], race [271], age [100] and others. As the Australian labor market is not immune to systemic biases [146], likely, the training data used for this research (HILDA – a representative household survey) reflects these systemic biases to some extent. Therefore, the results presented in this paper should be viewed as 'descriptive' of labor mobility in Australia rather than 'prescriptive' of individuals' career options. To help safeguard these systemic biases, we add a human-in-the-loop to filter the generated recommendations. The system we design in Fig. 6.5 is a decision-aid tool that filters top recommendations based on posting frequency (the link colors) to identify which occupations are growing in demand. Additional filters can be applied, such as top recommendations based on salary, education level, years of experience demanded, specific skill sets, industries, and others. While these filters do not remove biases from the recommendations entirely, they do provide individuals using this system with greater autonomy in exploring potential career paths.

6.4 Conclusion

Leveraging longitudinal datasets of real-time job ads and occupational transitions from a household survey, we have developed the SKILLS SPACE method to measure the distance between sets of skills. This enabled us to build systems that both *recommend* job transition pathways based on personalized skill sets and *detect* the growth of disruptive technologies in labor markets that could accelerate forced job transitions. Our *Job Transitions Recommender System* has the potential to assist workers, businesses and policy-makers to identify efficient transition pathways between occupations. These targeted and adaptive recommendations are particularly important during economic crises when labor displacement increases and workers are forced to transition to another job. The *Job Transitions Recommender System* could therefore assist with the current labor crisis caused by COVID-19. Additionally, it could assist with potential future crises, such as accelerated job transitions caused by AI labor automation.

We further demonstrate the usefulness and flexibility of SKILLS SPACE by applying it as a measure of AI adoption in labor markets. This acts as an 'early warning system' of forthcoming labor disruptions caused by the adoption and diffusion of AI within industries. Such a measure could complement other indicators of AI adoption, serving policy-makers and businesses to monitor the growth of AI technologies and its potential to accelerate job transitions.

While the future of work remains unclear, change is inevitable. New technologies, economic crises, and other factors will continue to shift labor demands causing workers to move between jobs. If labor transitions occur efficiently, significant productivity and equity benefits arise at all levels of the labor market [265]; if transitions are slow, or fail, significant costs are borne to both the State and the individual. Therefore, it is in the interests of workers, firms, and governments that labor transitions are efficient and effective. The methods and systems we put forward here could significantly improve the achievement of these goals.

6.5 Supplementary Materials

This section discusses the related works that have directly informed the research in SKILL-DRIVEN RECOMMENDATIONS FOR JOB TRANSITION PATHWAYS.

6.5.1 Using a standardized occupation taxonomy – ANZSCO.

All data sources mentioned above correspond to their respective occupational classes according to the Australian and New Zealand Standard Classification of Occupations (ANZSCO) [25]. ANZSCO provides a basis for the standardized collection, analysis and dissemination of occupational data for Australia and New Zealand. The structure of ANZSCO has five hierarchical levels - major group(1-digit), sub-major group (2-digit), minor group (3-digit), unit group (4-digit) and occupation group (6-digit). For visualizing the distance between occupations in Fig. 6.1B SKILL-DRIVEN RECOMMENDATIONS FOR JOB TRANSITION PATHWAYS, the 6-digit grouping level was applied. We used the 6-digit level here because (1) it is the most detailed occupational grouping and (2) the automation probabilities from Frey & Osborne were mapped at this level [151]. The presented results for the 'Job Transitions Recommender System' and the subset of the *Transitions Map* shown in Fig. 6.3 is grouped at the 4-digit unit grouping level. Occupational recommendations were made at this level to match the ground-truth of actual transitions from the Household, Income and Labour Dynamics in Australia (HILDA) longitudinal dataset [130].

Shortcomings of ANZSCO. There are shortcomings to analyzing occupations within ANZSCO classifications. Official occupational classifications, like ANZSCO, are often static taxonomies and are rarely updated. They therefore fail to capture and adapt to emerging skills, which can misrepresent the true labor dynamics of particular jobs. For example, a 'Data Scientist' is a relatively new occupation that has not yet received its own ANZSCO classification. Instead, it is classified as an 'ICT Business & Systems Analyst' by ANZSCO, grouped with other job titles like 'Data Analysts', 'Data Engineers', and 'IT Business Analysts'. However, as ANZSCO is the official and prevailing occupational classification system, all data used for this research are in accordance with the ANZSCO standards.

Mapping O*NET to ANZSCO. To leverage the strengths of earlier research by Frey and Osborne [151] on the occupational risks of labor automation caused by AI technologies, we first needed to map O*NET occupations to ANZSCO, so as to take advantage of their automation risk probabilities at the 6-digit level. O*NET is a standardized and publicly available database of labor market data in the United States [326]. The occupations, however, are slightly different compared to ANZSCO. Therefore, we used a concordance table from the Australian Federal Department of Education, Skills and Employment [37] to map O*NET occupations to ANZSCO at the 6-digit level. This resulted in each ANZSCO occupation being assigned an automation risk probability according to the Frey and Osborne research.

6.5.2 Model Features

Below is a summary of the features included within the 'Job Transitions Recommender System' models. The features have been grouped into 'Labor Demand' (job ads data) and 'Labor Supply' (employment statistics) categories for ease of review. Each feature in 6.1 is measured at the ANZSCO 4-digit level per calendar year from 2012-2018. This reflects the first available year of the longitudinal job ads data (2012) and the most recent year of the 'ground-truth' HILDA data (2018). The 'source' and 'target' occupations are independently associated with each of the features in 6.1. However, the 'theta' (distance between skill sets measure) and 'Difference' features relate to both the 'source' and 'target' occupational pair. In total there are 19 features.

6.5.3 Validation

Statistical Test. To obtain initial validation of the SKILLS SPACE distance measures, we conducted a paired statistical test, as explained in SKILL-DRIVEN RECOMMENDATIONS FOR JOB TRANSITION PATHWAYS. To run this experiment, we labeled each 'source' and 'target' occupational pair with their distance measure for their given year (called the 'True Sample'). We then simulated an alternate sample of transitions where we maintain the same 'source' occupations and randomly select 'target' occupations, all assigned with their pairwise distance scores (called 'Simulated Sample'). Section 6.5.3 shows the distribution of all job transitions, including transitions to the same occupation. We found that the differences between the 'True' and 'Simulated' transition samples are statistically significant (t-statistic = 16.272, p-value = 2.706911e-58, Cohen's D effect size = 0.42). However, 909 of the 2999 (or 30%) of the yearly job transitions from 2012-2018 are movements to the same occupation. Intuitively, the skill set distance of transitioning to another job in the same occupation is likely small, especially compared to other occupations. Therefore, we wanted to test whether the statistical significance holds when we exclude transitions to the same occupation. To run this test, we first

Table 6.1: Summary of constructed features and their explanation.

	Feature	Description
Labor Demand	theta:	SKILLS SPACE distance between 'source' and 'target' occupation
	Posting Frequency:	number of job advertisement vacancies
	Posting Frequency Difference:	difference between the 'source' and 'target' posting frequency
	Median Salary:	maximum median salary advertised
	Salary Difference:	difference between the 'source' and 'target' salaries
	Minimum Education:	minimum years of formal education required
	Education Difference:	difference between the 'source' and 'target' years of formal education required
	Minimum Experience:	minimum years of formal experience required
	Experience Difference:	difference between the 'source' and 'target' years of experience required
Labour Supply	Total Employed:	total number employed at ANZSCO Unit level (000's)
	Total Employed Difference:	difference between the 'source' and 'target' of total number employed at ANZSCO Unit level (000's)
	Total Hours Worked:	total hours worked at ANZSCO Unit level (000's)
	Total Hours Worked Difference:	difference between the 'source' and 'target' of total hours worked at ANZSCO Unit level (000's)

removed job transitions to the same occupation (leaving 2090 occupations from 2012-2018). Following the same process described above, we created 'True' and 'Simulated' samples. Section 6.5.3 illustrates the differences between the two samples, which are again statistically significant (t-statistic = 4.514, p-value = 6.534642e-06); however, the effect size is lowered (Cohen's D effect size = 0.14).

We repeated the procedure 100 times: we generated 100 'Random' populations and we perform the statistical testing for each. As shown in SKILL-DRIVEN RECOMMENDATIONS FOR JOB TRANSITION PATHWAYS, 87 of the 100 obtained p-values were lower than 0.05 providing confidence in the statistical significance of SKILLS SPACE to represent occupational transitions in this research.

Job Transitions Recommender System validation. Fig. 6.8 shows the confusion matrix for the binary classifier model containing all of the features from Table 6.1. This feature configuration achieved the highest performance (Accuracy = 76% and F1 Macro Average = 77%). As observed, this trained model was able to predict *True Negatives* ('Not a Transition' – Recall = 84%) slightly better than *True Positives* ('Actual Transition' – Recall = 71%). Fig. 6.9 shows the 'receiver operating characteristic' curve (ROC curve), which is

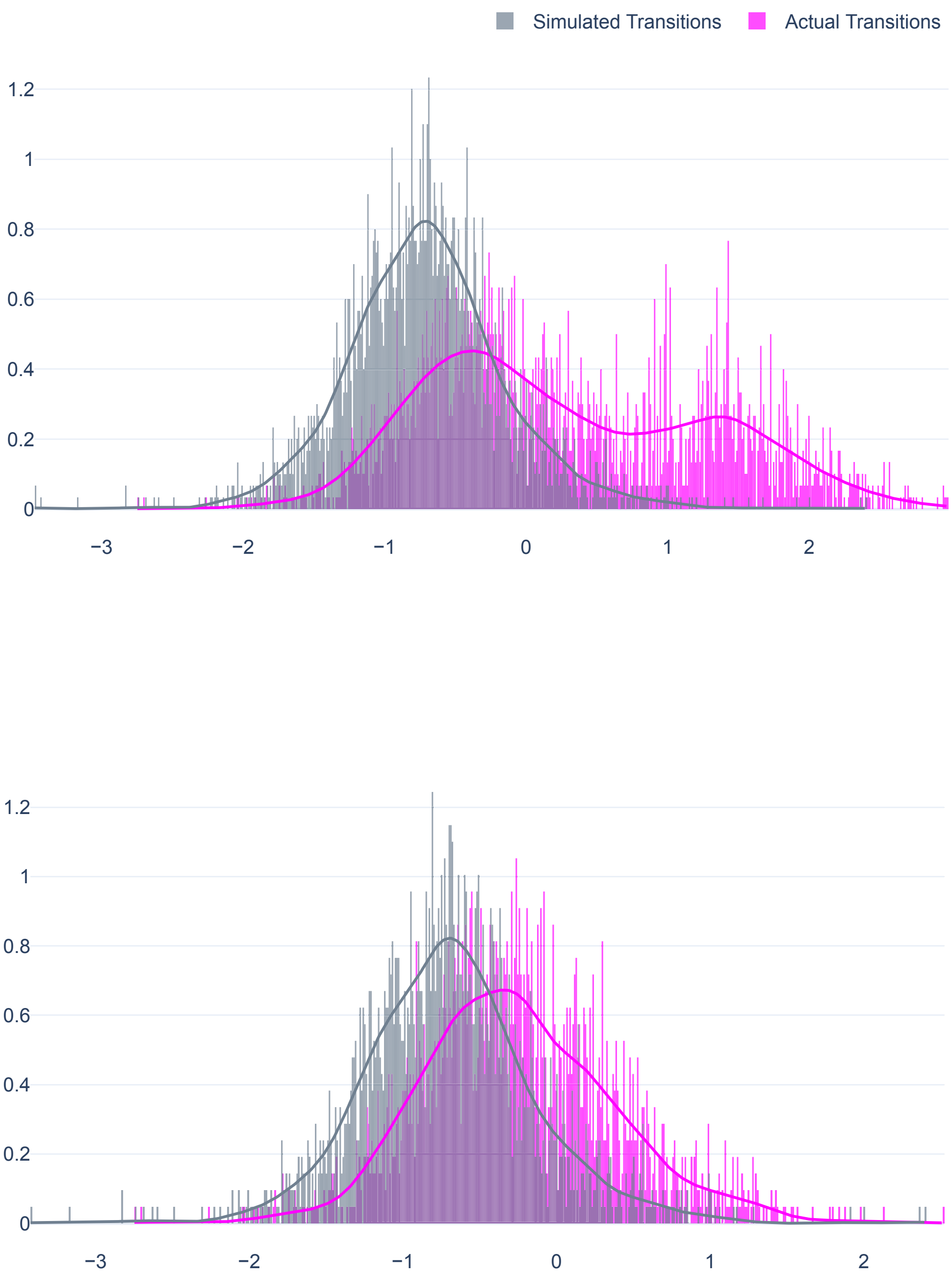


Figure 6.7: (a) Statistical test against all occupational transitions and (b) against transitions where the worker changed occupations

the performance of the binary classification model at all classification thresholds. ROC curves summarize the trade-off between the *True Positive* rate (y-axis) and *False Positive* rate (x-axis) for the classifier model using different probability thresholds. Generally, high-performing models are represented by ROC curves that bow up to the top left of the plot. As illustrated in Fig. 6.9, the blue ROC curve is consistently above the diagonal red dashed line that represents a 50% probability – models that perform below this dashed line are no better than random chance. This reinforces that the model consisting of 'All Features' achieves strong performance levels.

Similarly, Fig. 6.10 shows the confusion matrix and Fig. 6.11 illustrates the ROC curve for the classifier model that only includes the SKILLS SPACE distance measure ('theta'). While the 'theta only' model still performs relatively well (Accuracy & F1 Macro Average = 73%), performance does decline. Again, the *True Negatives* ('Not a Transition' – Recall = 83%) outperform the *True Positives* ('Actual Transition' – Recall = 64%). This highlights that the added labor market features from job ads data and employment statistics increased the performance capabilities of the models to predict *True Positives*. These performance differences are also represented in Fig. 6.11 showing a slightly lower ROC curve for models that used 'theta' alone.

Ablation Test and Feature Importance. In order to understand the relative importance of the modeled features in the 'Job Transitions Recommender System', we conducted an ablation test and feature importance analysis. An ablation test involves iteratively removing one feature from the feature set and then re-training the model to make predictions and evaluate performance. We conclude that a feature is 'more important' to a model's predictive capabilities if performance declines when it is removed. Section 6.5.3 shows the results of all 19 features, highlighting that the largest performance decline occurred when the 'theta' distance measure was removed. These models were all trained with a consistent setup, as explained in SKILL-DRIVEN RECOMMENDATIONS FOR JOB TRANSITION PATHWAYS.

To reinforce the results from the ablation test, we then conducted a feature importance analysis as seen below. We use the 'Gain' metric, which shows the relative contribution of each feature to the model by calculating the features' contribution for each tree in the XGBoost model. A higher gain score indicates that a feature is more important for generating a prediction. Again, 'theta' is overwhelmingly identified as the most important feature for predicting job transitions.

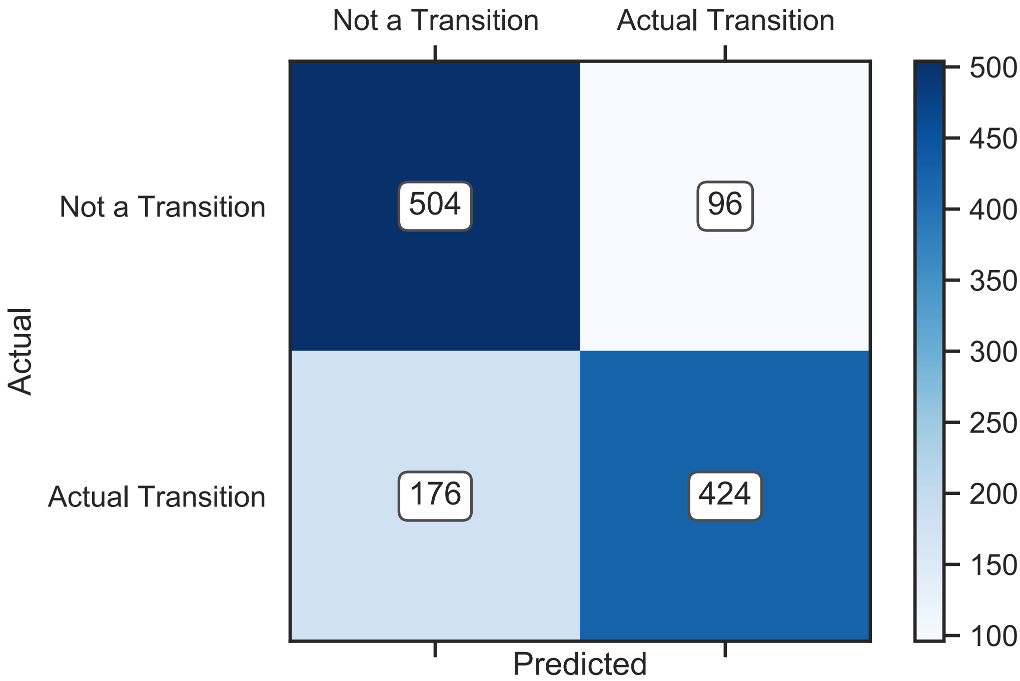


Figure 6.8: Confusion matrix with all features included.

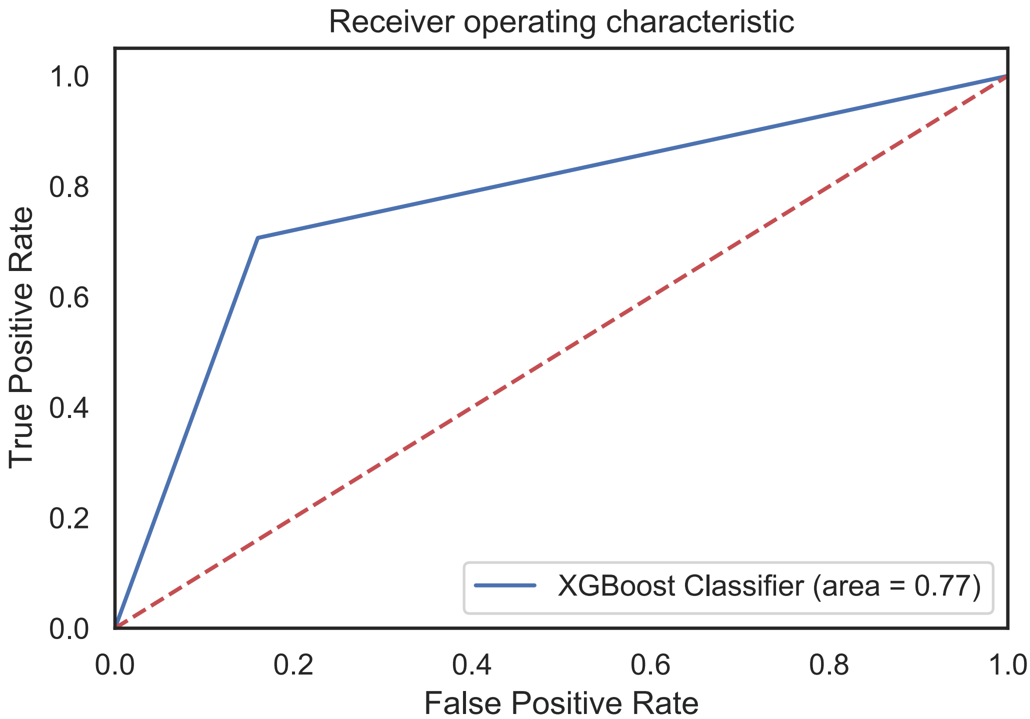


Figure 6.9: ROC curve with all features included.

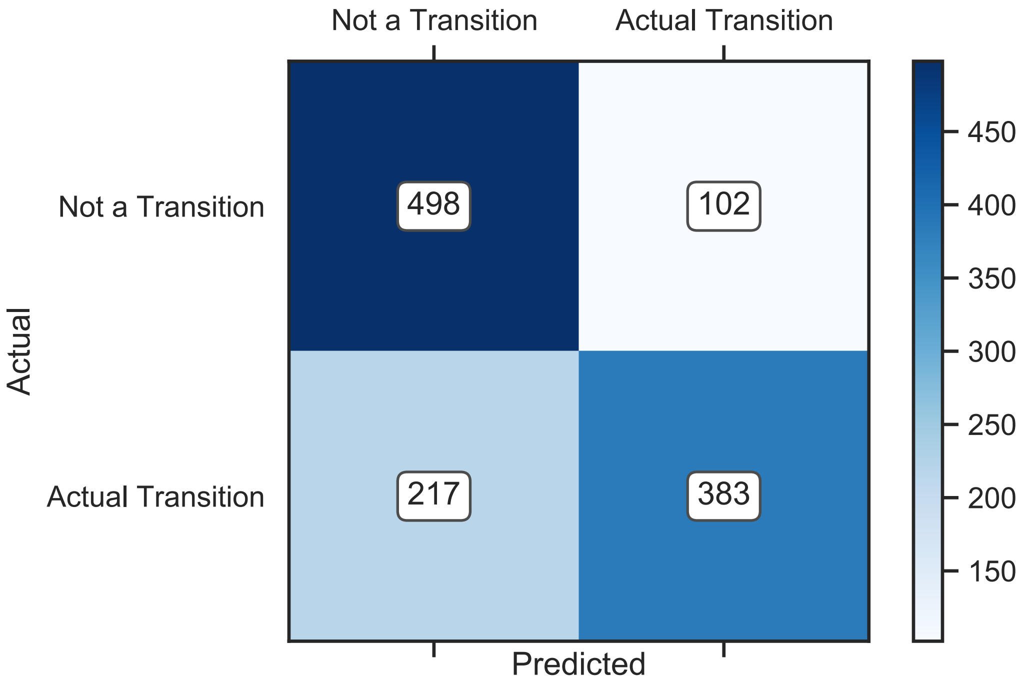


Figure 6.10: Confusion matrix for theta only.

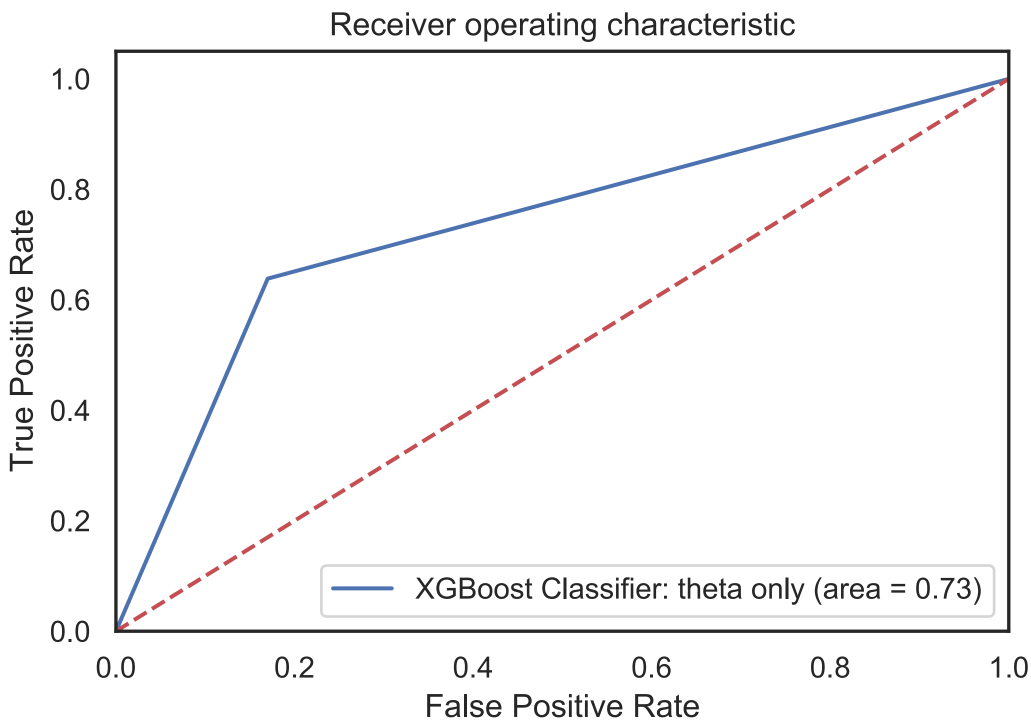


Figure 6.11: ROC curve with theta only.

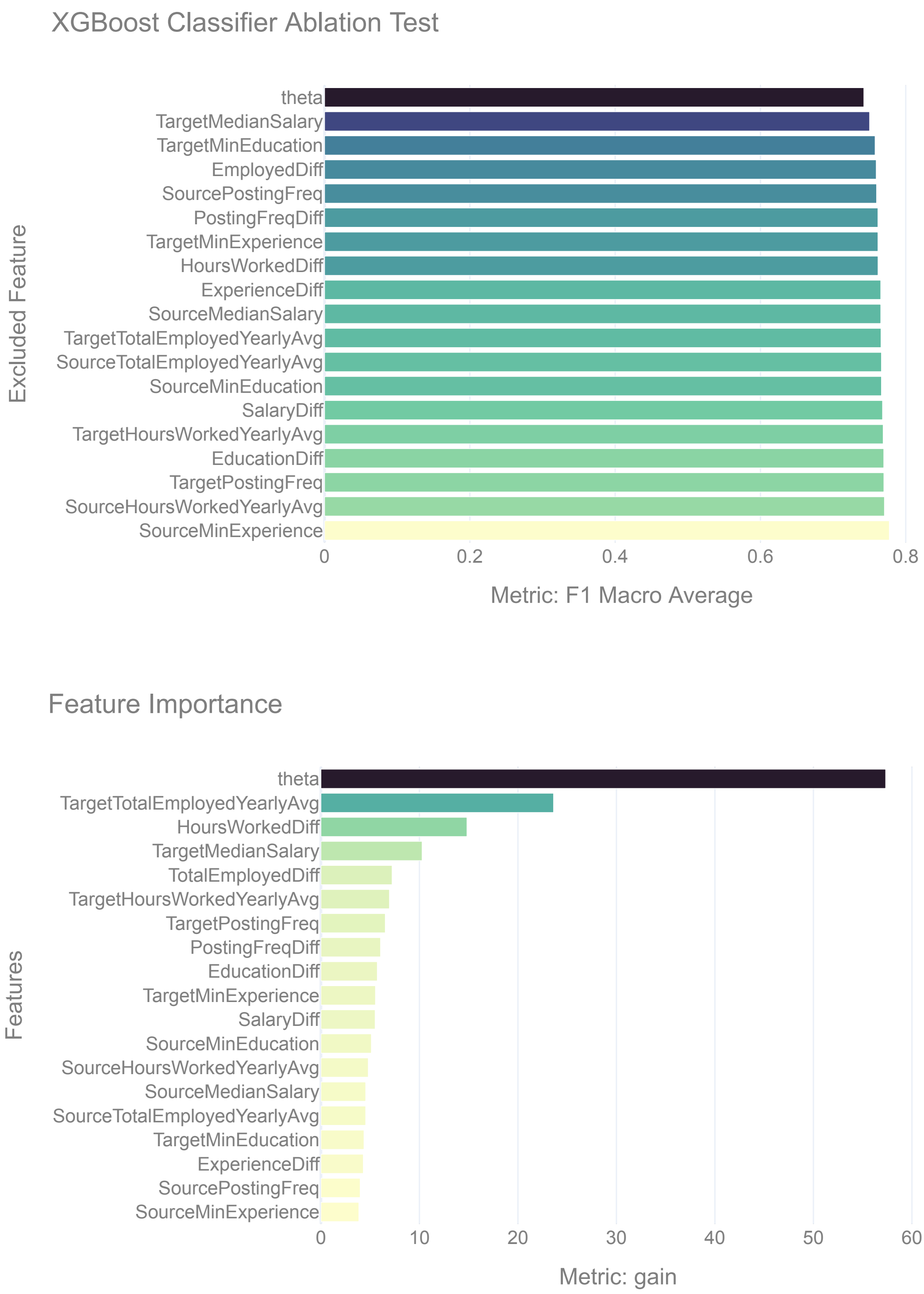


Figure 6.12: (a) Ablation test of classifier features and (b) feature importance analysis

6.5.4 Recommending Jobs & Skills

Transitions Example – Domestic Cleaner

Occupation	Transition Probability	Num. Job Ads 2019	Num. Job Ads 2020	Difference	Percentage Difference
Domestic Cleaners	0.960395	323	276	-47	-14.551084
Commercial Cleaners	0.946621	865	671	-194	-22.427746
Waiters	0.943874	690	253	-437	-63.333333
Bar Attendants and Baristas	0.937961	600	180	-420	-70.000000
Sales Assistants (General)	0.935315	2835	1609	-1226	-43.245150
Chefs	0.926472	1904	877	-1027	-53.939076
Cooks	0.914349	726	356	-370	-50.964187
Aged and Disabled Carers	0.893725	961	1302	341	35.483871
Child Carers	0.887601	837	414	-423	-50.537634
General Clerks	0.876921	2281	1466	-815	-35.729943

6.5.5 AI Adoption

AI Similarity Scores. The table below contains the underlying data for the AI Adoption radar chart in the *Developing a Leading Indicator of AI Adoption* section.

Industry	2013	2016	2019	Percentage change 13-19
Financial and Insurance Services	0.000958	0.001599	0.002887	201.395294
Information Media and Telecommunications	0.001057	0.001283	0.002286	116.285278
Professional, Scientific and Technical Services	0.000545	0.001027	0.001590	191.537693
Retail Trade	0.000266	0.000568	0.001348	407.375732
Electricity, Gas, Water and Waste Services	0.000443	0.000520	0.001300	193.594618
Education and Training	0.000707	0.000933	0.001257	77.744881
Transport, Postal and Warehousing	0.000282	0.000427	0.000981	247.957226
Public Admin	0.000270	0.000481	0.000905	234.594611
Rental, Hiring and Real Estate Services	0.000157	0.000439	0.000775	392.117595
Arts and Recreation Services	0.000256	0.000571	0.000738	188.519596
Manufacturing	0.000206	0.000356	0.000690	235.589203
Administrative and Support Services	0.000243	0.000346	0.000661	172.344264
Wholesale Trade	0.000219	0.000388	0.000642	193.307221
Mining	0.000260	0.000223	0.000595	128.900314
Construction	0.000136	0.000236	0.000487	257.990332
Other Services	0.000136	0.000180	0.000306	124.906123
Health Care and Social Assistance	0.000080	0.000170	0.000281	252.858476
Accommodation and Food Services	0.000086	0.000208	0.000267	208.987030
Agriculture, Forestry and Fishing	0.000108	0.000183	0.000239	120.979250

Temporal AI skill similarity to Australian Industries. The figure below is another visualization of the same data in the *Developing a Leading Indicator of AI Adoption* section illustrating all of the yearly AI similarity scores for the Industries from 2012-2019.

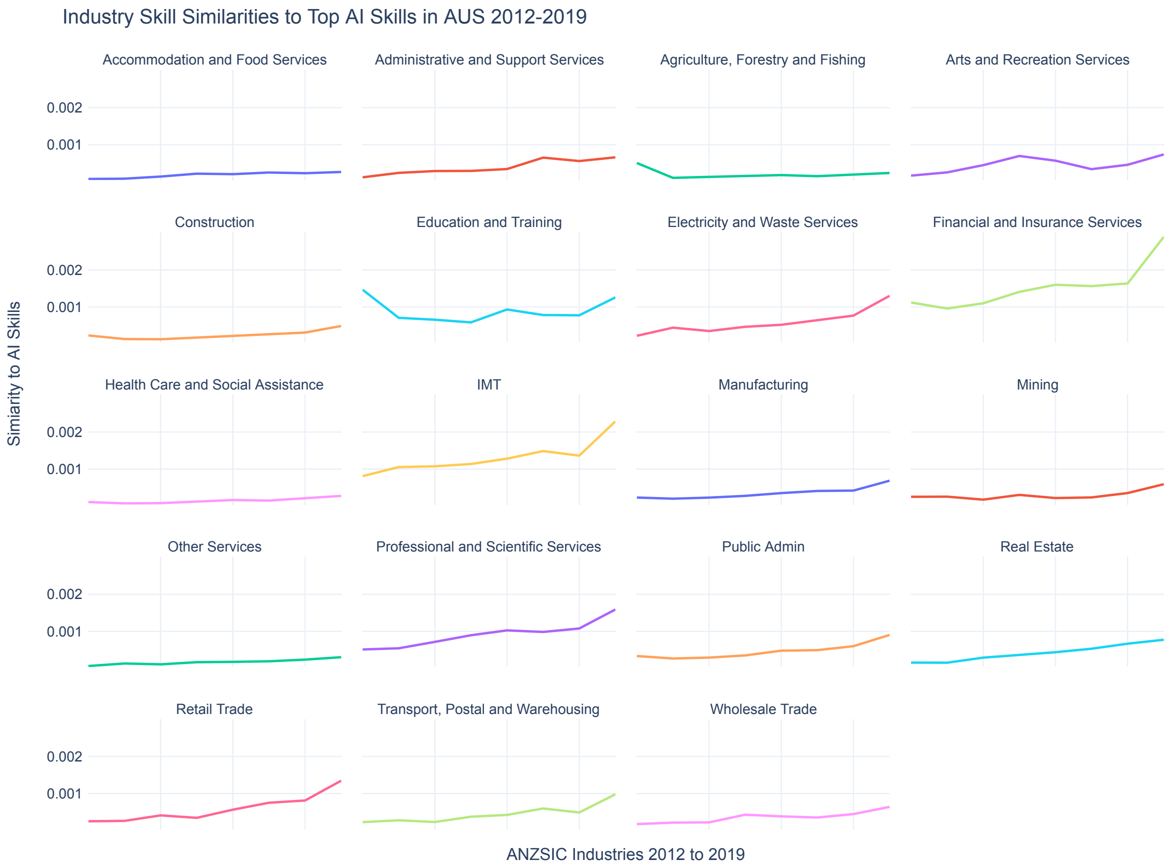


Figure 6.13: Yearly skill similarities between AI skills and Australian Industry (ANZSIC Division) skill sets from 2012-2019.

6.5.6 Skill Count Distribution

As discussed in *Materials & Methods*, before calculating individual skill similarities, we filtered out extremely rare skills to reduce noise and computational complexity. We set the minimum yearly skill count threshold to be greater than or equal to 5.

As seen in the Empirical Cumulative Distribution Function (ECDF) in Section 6.5.6, this threshold represents over 75% of all skills in 2018 (6,981 skills). All skills on the left side of the dotted threshold line were excluded, which accounted for <25% of skills.

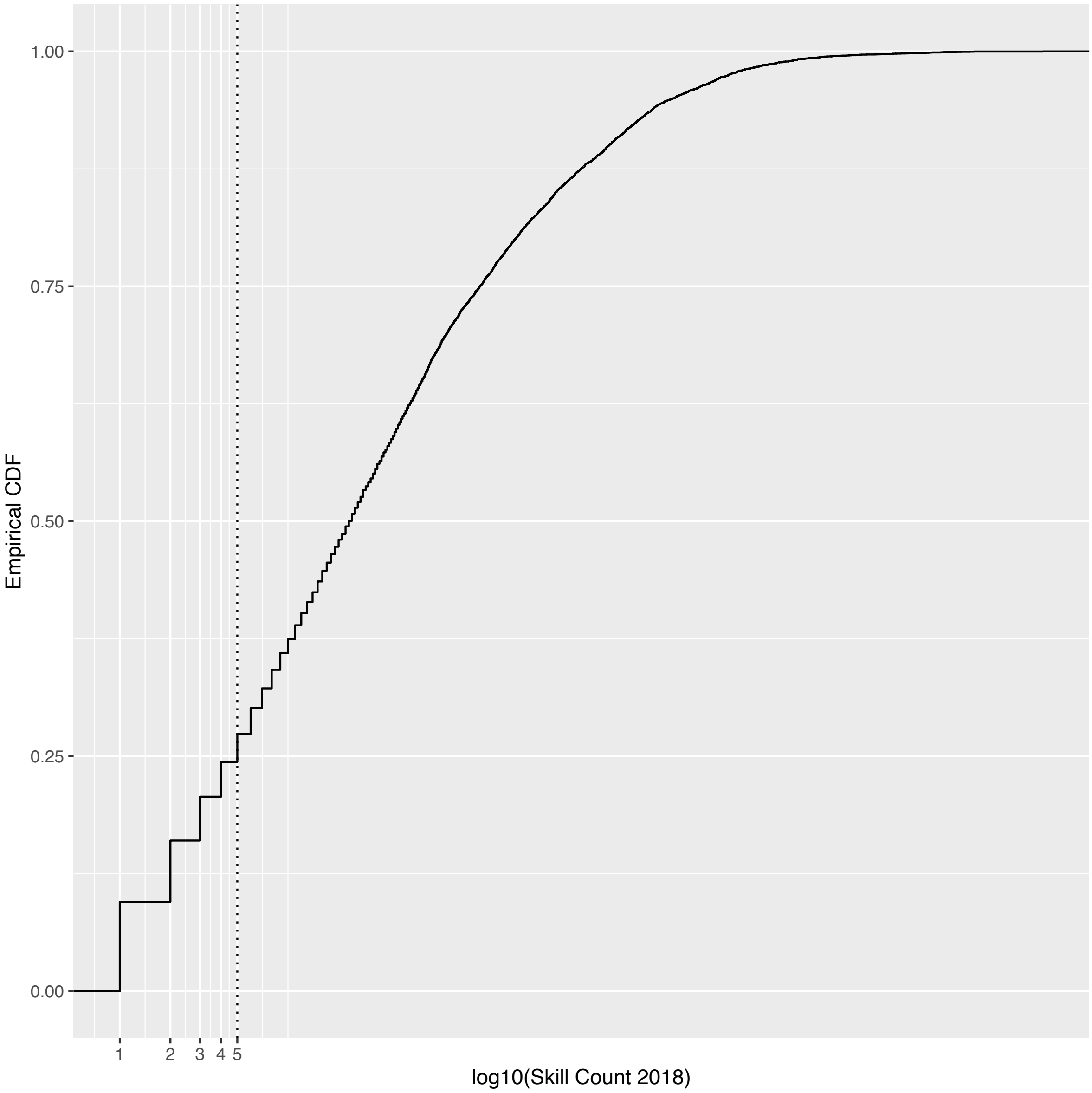


Figure 6.14: Empirical Cumulative Distribution Function of skill counts within job ads for 2018.

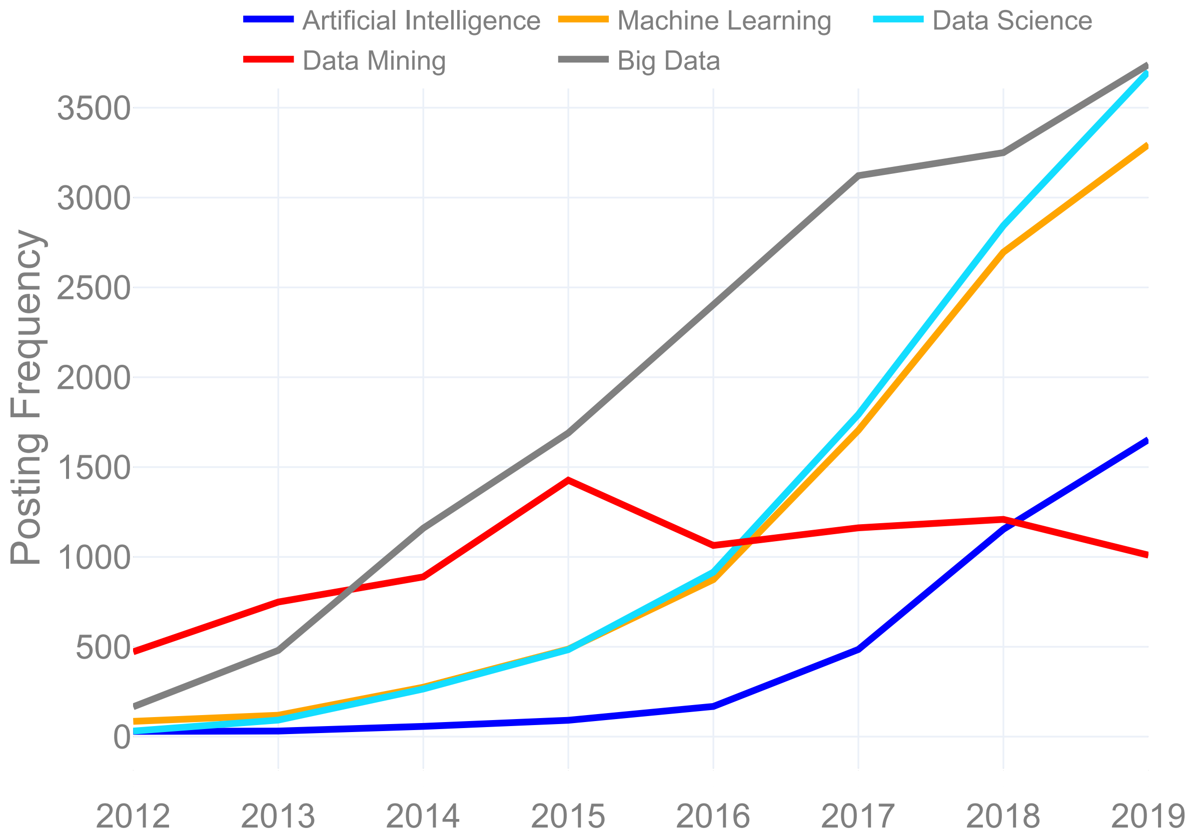


Figure 6.15: Yearly posting frequency of the five AI seed skills used to build a dynamic list of yearly AI skills.

6.5.7 Posting Frequency of AI Seed Skills

We selected five 'seed skills' to construct a dynamic list of yearly AI skills, as described in Section 6.2 and visualized in Fig. 6.6. This allowed us to measure the distance between AI skills and industry skill sets, capturing both the evolution of skill demands and accounting for skill importance. The most common method, however, is to simply count the frequency of a skill (or group of skills) over time. In Section 6.5.7, we show the posting frequency of the five AI seed skills used to construct the dynamic list of yearly AI skills. The five AI seed skills being: (1) Artificial Intelligence; (2) Machine Learning; (3) Data Science; (4) Data Mining; and (5) Big Data.

As Section 6.5.7 shows, the posting frequency for all five seed skills increases from 2012 to 2019, albeit at different rates. 'Data Science' experiences the steepest increases over this period. Whereas 'Data Mining' has had more modest growth, reaching its highest posting frequency levels in 2015 and has since declined.

Section 6.5.7 shows that not only have the absolute posting frequencies of the AI seed

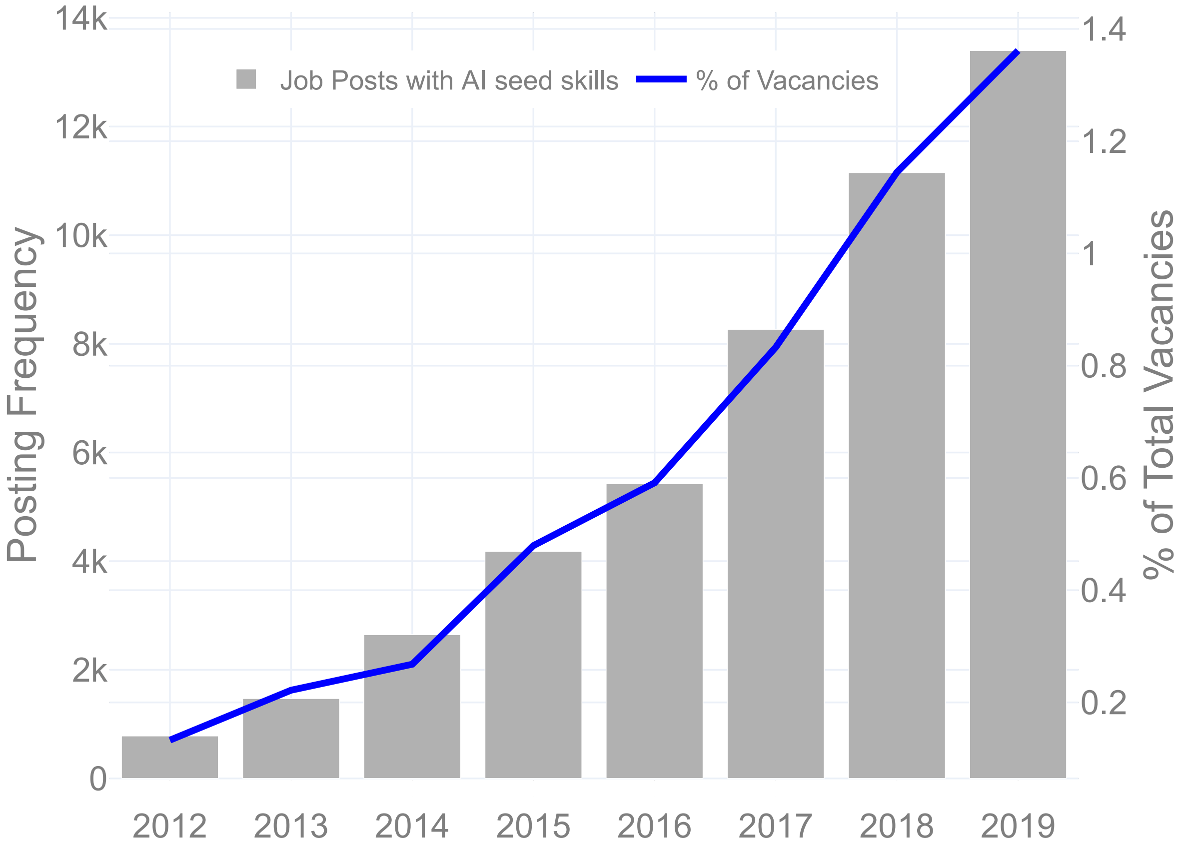


Figure 6.16: The percentage of vacancies in Australia that contain these five AI seed skills.

skills increased, but also the percentage of vacancies containing these skills. In 2012, approximately 0.13% of Australian job ads (or 783 vacancies) contained at least one of the AI seed skills. In 2019, this had increased to more than 1.3% of job ads (or 13,399 vacancies) – growth of over ten times the percentage of job ads requiring the AI seed skills.

While these simple metrics provide an indication of the degree of growth required by the AI seed skills, there are some fundamental shortcomings for using posting frequencies as a proxy for labor demand. A discussion on the disadvantages of using posting frequency compared to skill similarity follows.

6.5.8 Advantages of Skill Similarity over Posting Frequency

The proxy most widely used in literature [101] for skill importance is skill frequency. This simply counts how many times a skill appears in job ads associated with a given occupation (or other groups) during a predetermined period of time; the higher the count, the greater the demand and, implicitly, the greater importance of the skill to the occupation. While skill frequency can provide some indication of labor demand, it fails to normalize for skills that are demanded by all or most jobs. This does not necessarily reveal which skills are more or less important to a given occupation, as some skills generalize across all occupations at high frequencies. For example, 'Communication Skills' and 'Teamwork' occur in over one-quarter of all job ads and are ubiquitous across all occupations). However, we know that some skills are more important than others to specific jobs. We therefore capture a proxy for skill importance by measuring the comparative advantage of each skill in each job ad, as seen in the RCA equation in *Materials & Methods*. Our measure controls for high-occurring skills through normalization and develops a measure of skill importance within individual job ads that later represent skill importance within labor market groups (occupations, industries etc.).

Chapter 7

Discussion & Conclusion

The nature of work continues to change in Australia [111, 219, 329]. Skills and tasks within jobs are constantly evolving, altering the fabric of the labour market. The demand for and supply of labour dynamically adjusts to a range of socio-economic factors [111], of which technological change is among the most significant [7, 68, 182]. With change comes adjustment periods. Structural adjustments to labour markets can cause problematic issues, such as skill shortages and forced job transitions. These market inefficiencies can result in unfavourable outcomes for individuals, firms, and macroeconomies. While these issues brought about by structural changes in labour markets are not new, the rapid advances of data science and machine learning methods present research opportunities to these established problems of labour economics.

This thesis has presented a series of papers that apply data-driven methods to major issues relating to the changing labour market dynamics in Australia. The compilation of papers included in this dissertation address the specific problems of skill shortages, job transitions, and AI technology adoption. The methods, models, and systems I have developed for this doctoral research have direct application to policy-makers, businesses, and job-seekers. Whether it is more accurately predicting occupational skill shortages in advance, dynamically detecting the growth of AI skills as a proxy for AI adoption, or recommending jobs and skills for workers faced with the challenges of transitioning to another career, this research offers data-driven approaches to major and prescient labour market problems. Arguably, the importance of these works have been elevated by the economic and labour market crises induced by COVID-19. With unemployment rates in Australia reaching their highest levels in over two decades toward the beginning of the pandemic (see Fig. 2.1) and sweeping restrictions imposed to entire industries, many workers were forced to change their careers and find new jobs. The 'Job Transitions Recommender System', presented in Chapter 6, provides a tool that can help workers navigate this often distressing challenge.

7.1 Limitations and future research

The remainder of this Chapter will consist of a discussion on the limitations of this work and future research opportunities. This will be organised within the three specific focus areas of this thesis: skill shortages, job transitions, and AI adoption.

7.1.1 Skill shortages

The main limitations of the research on skill shortages in Chapter 3 and Chapter 4 are their reliance on previous research and datasets for classifying occupational skill shortages. Chapter 3 leveraged previous works that found 'Data Science and Analytics' occupations to be in shortage in Australia [123] and other labour markets similar to Australia [70, 227]. Whereas Chapter 4 used the 'Historical list of skill shortages in Australia' from the Australian Federal Department of Education, Skills and Employment (DESE) [124], which longitudinally tracks the skill shortage status of 132 occupations, to construct a ground-truth dataset to train machine learning classifiers against. While these sources are sound, they are not without their limitations. First, as relatively new occupations, such as 'Data Scientists' and 'Data Engineers', are not represented in ANZSCO, it is difficult to verify the extent of labour shortages for these occupations. Second, for the occupations where skill shortages are longitudinally measured [124], they depend on lagging survey results and cover a non-representative sample of occupational classes in Australia. Therefore, constructing alternate datasets to accurately measure occupational skill shortages represents a major research opportunity for future work.

One avenue for constructing a **representative and timely dataset on occupational skill shortages** could be using proxy measures of labour demand and labour supply to calculate normalised disequilibria. That is, developing measures to identify when labour demand for occupations exceed the labour supply of occupations in the job market, or the inverse scenario. These could provide timely measures of occupational skill mismatches. As previously described in Chapters 2-4, occupational skill shortages occur when the demand for an occupation exceeds the supply of labour available at a prevailing market rate; conversely, skill surpluses emerge when the supply of occupations exceed the demand for workers that possess particular skill sets [284]. As the majority of employment vacancies in Australia are now advertised online [126], job ads data are among the leading measures of labour demand. Combining jobs ads data with occupational employment statistics could provide the basis for measuring disequilibria between labour demand and labour supply of occupations. This approach could yield an occupational skills mismatch dataset that both represents a greater number of occupations measured and regularly updates (likely quarterly in accordance with the ABS Labour Force Detailed Quarterly release [35]). However, a number of prerequisites would need to be met and issues resolved. First, this work assumes that the researcher will have access to job ads data that is sufficiently large (for example, one of the major job boards) and classified into their respective ANZSCO occupational classes. Second, the labour demand and supply levels will need to be normalised to establish disequilibria thresholds. And third, the researcher will need to account for the time lags associated with employment statistics. Job ads data are real-time indicators of employer demands, whereas employment statistics are lagging indicators of employment levels from previous quarters. Accounting for these timing differences will be an essential consideration for constructing such a dataset.

Another research opportunity regarding skill shortages is the **detection of emerging skills and occupations**. As shown in Chapter 3 and Chapter 5, new skills and occupations emerge and fade away over time. For example, the emergence and growth of ‘Data Scientists’ since 2012 (see Chapter 3) and the continued rise of ‘Social Media’ related skills in Australia (see Chapter 5). *Ex post* these developments can appear obvious and logical. However, detecting such developments at the time, or predicting in advance, is extremely difficult. Currently, this is an under-served area of research. It also represents a significant opportunity for the topic of skill shortages, as the early detection of emerging skills and occupations can facilitate targeted interventions, such as education and training programs to develop emerging skills and help to prevent shortages. Giabelli et al. [160] have made recent progress on this open problem. The authors have developed a system that automatically updates the European Skills, Competences, Qualifications and Occupations (ESCO - the European standardised skill and occupational taxonomy) with new occupation terms extracted from online job ads. The system, called ‘NEO’, measures the pairwise semantic similarity between words in a taxonomy to suggest emerging occupations from job ads and their relatedness to existing occupational terms [160]. There are further research opportunities at the skill-level. Specifically, measuring the emergence and disappearance of skills using job ads data. One such line of inquiry could involve temporal clustering of skills to observe how the composition of skill clusters change over time. As new skills emerge, and old skills disappear, the characteristics of skill clusters will evolve. The ability to observe these changes to skill clusters could assist with the early detection of emerging skills, which could also help to detect emerging occupations.

7.1.2 Job transitions

While the methods and systems developed for making job transition recommendations in this thesis are novel contributions, there are limitations and notable avenues for future work. The job transitions recommender system presented in Chapter 6 identifies occupational transition pathways and top skills to acquire at the aggregated occupational level. That is, it does not differentiate for differences *within* an occupation. Developing a system to make **personalised recommendations** represents a significant next step in the evolution of this research. This would allow for personalised constraints to be implemented, such as targeting locations, preferred occupations, salary ranges, education levels, years of experience, and the selection of skills an individual has already acquired. Further, there is potential to extend the application of this **recommender system for groups of people**, for purposes such as reskilling organisations. Scaling the recommender system to groups of people creates an additional layer of complexity, as recommendations are no longer optimised at the individual level but for groups of individuals. This raises the established optimisation problem in Computer Science where the local optimum does not always result in the global optimum - or, what is best for the individual is not necessarily best for the group.

Another limitation of these works on job transitions is that they exclusively use labour market information. However, we know that there are other factors that influence job decisions, such as career goals, personality preferences, and lifestyle factors [195, 293]. Therefore, finding ways to incorporate such information into job transition models is an exciting area for future research. One such opportunity is **incorporating personality profiles** as explanatory features in addition to labour market information. Recently, researchers were able to accurately predict occupations according to personality traits collected from social media activity [202]. This research found that Schwartz's 'Big 5' personality traits are predictive of occupational decisions. Developing ethical and unobtrusive ways to incorporate personality profiles with detailed occupational and skills data could significantly improve job transition recommendations.

Last, the size of the job transitions ground-truth dataset used for training the models in Chapter 6 was also a limitation that can improved upon. While the quality of the HILDA dataset is high [130], it consisted of only 2,999 job transitions from 2012-2018. A future research opportunity, therefore, is to acquire a larger ground-truth dataset of job transitions to train against. This could come from sources such as LinkedIn or Taleo profiles. There is also potential to combine multiple public datasets from different labour markets that share similar demographics, such as Canada, the United Kingdom and the United States. The opportunities for research on job transitions are immense and will likely grow in importance over the forthcoming decade as the labour markets around the world attempt to rebuild from the economic damage caused by COVID-19.

7.1.3 AI adoption

AI technology adoption at the firm-level depends on a range of factors, as discussed in Chapter 2. Access to skilled labour for firms to implement and make productive use of these technologies represents one of these adoption factors. Using skills data enabled this thesis to observe how the demand for AI-related skills and occupations have changed over time, which were developed into leading indicators of AI adoption. As previously stated, however, the availability of skilled labour represents one determinant of AI adoption, which is a limitation of this research. Future work could build upon this research by **triangulating these results with other indicators of AI adoption**, such as firm-level survey results, increases in computing spend by firms, patents, acquisitions of AI-related companies, and others.

Another significant area of future work is accurately **measuring the effects of AI on productivity**. For many General Purpose Technologies (GPTs) that have transformed economies and labour markets, such as electricity and personal computers, there were extended transition periods of perceived stagnation that required significant complementary investments [9, 15, 80]. Firms needed to create new business processes, products, business models, and hire skilled labour to make productive use of these technologies. This often involved intangible investments that are not readily measured in productivity statistics. As Robert Solow famously quipped in 1987 about the personal computing revolution: *"You can see the computer age everywhere but in the productivity statistics"*. Brynjolfsson et al. [90] argue that 'Solow's paradox' is a more general phenomenon that has applied to a number of historical GPTs and could be occurring today with AI. The authors develop a model that can explain the underestimation of productivity growth during the advent of GPTs before the intangible investments are realised and total factor productivity surges – the authors call this the 'Productivity J-Curve' [90]. Measuring the impacts of emerging technologies, such as AI, on productivity statistics represents a rich area of research that is closely related to studies on AI adoption.

7.1.4 A final note

The long arc of changes to Australia's labour market have bent towards progress [219]. In recent decades, unemployment has remained relatively low, labour participation has increased, and educational attainment has been high creating a skilled and dynamic labour force [111, 182]. This is not to dismiss the challenges that structural changes pose to labour markets, such as adjusting to emerging technologies or responding to exogenous shocks, like COVID-19. Nor is it to place blind faith in the determinism of history. The challenges of transitioning displaced workers to meet the new labour demands are likely to be significant. But the demand for labour is unlikely to be the primary concern.

Instead, the concern should be focused on equipping workers with the skills to quickly transition to meet new labour demands. In other words, it is likely that old jobs will be replaced by new jobs, but it is the speed of transitioning workers to these new jobs that could be problematic.

A large part of this transition speed will depend on the types of jobs created and the skills demanded to fulfil these jobs. As economist James Bessen sums up, the problem is *"scarce skills not scarce jobs."* [68]

Data-driven approaches to the established research problems of labour economics represents a young and exciting research field. The works presented in this thesis contributes data-driven research to the issues of skill shortages, job transitions, and AI adoption in the changing Australian labour market from 2012-2020.

Bibliography

[1] *Introduction to boosted trees — xgboost 1.2.0-SNAPSHOT documentation*. https://xgboost.readthedocs.io/en/latest/tutorials/model.html. Accessed: 2020-8-5.

[2] R. Abbott and B. Bogenschneider, *Should Robots Pay Taxes? Tax Policy in the Age of Automation*, Harvard Law and Policy Review, 12 (2017), pp. 145–175.

[3] J. H. Abbring, G. J. van den Berg, P. A. Gautier, A. G. C. van Lomwel, J. C. van Ours, and C. J. Ruhm, *Displaced Workers in the United States and the Netherlands*, in Losing Work, Moving on: International Perspectives on Worker Displacement, P. J. Kuhn, ed., W.E. Upjohn Institute for Employment Research, July 2002.

[4] ABC, *Australia's social distancing rules have been enhanced to slow coronavirus — here's how they work*, ABC News, (2020).

[5] P. M. Abernathy, *The Expanding News Desert*, tech. rep., University of North Carolina at Chapel Hill, 2018.

[6] ACCC, *Digital platforms inquiry - final report*, tech. rep., Australian Competition and Consumer Commission, July 2019.

[7] D. Acemoglu and D. Autor, *Skills, Tasks and Technologies: Implications for Employment and Earnings*, in Handbook of Labor Economics, D. Card and O. Ashenfelter, eds., vol. 4, Elsevier, Jan. 2011, pp. 1043–1171.

[8] D. Acemoglu and P. Restrepo, *Artificial Intelligence, Automation and Work*, Tech. Rep. 24196, the National Bureau of Economic Research, Jan. 2018.

[9] D. Acemoglu and J. A. Robinson, *Why nations fail: The origins of power, prosperity, and poverty*, Currency, 2012.

[10] P. AGHION AND P. HOWITT, *On the Macroeconomic Effects of Major Technological Change*, Annales d'Économie et de Statistique, (1998), pp. 53–75.

[11] P. AGHION, B. F. JONES, AND C. I. JONES, *Artificial Intelligence and Economic Growth*, National Bureau of Economic Research, (2017).

[12] A. AGRAWAL, J. GANS, AND A. GOLDFARB, *Economic Policy for Artificial Intelligence*, in Innovation Policy and the Economy, Volume 19, University of Chicago Press, May 2018.

[13] ——, *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, Apr. 2018.

[14] A. ALABDULKAREEM, M. R. FRANK, L. SUN, B. ALSHEBLI, C. HIDALGO, AND I. RAHWAN, *Unpacking the polarization of workplace skills*, Science Advances, 4 (2018), p. eaao6030.

[15] R. C. ALLEN, *Engels' pause: Technical change, capital accumulation, and inequality in the british industrial revolutio*, Explorations in Economic History, 46 (2009), pp. 418–435.

[16] E. AMID AND M. K. WARMUTH, *TriMap: Large-scale Dimensionality Reduction Using Triplets*, ArXiv e-prints, (2019).

[17] D. AMODEI, C. OLAH, J. STEINHARDT, P. CHRISTIANO, J. SCHULMAN, AND D. MANÉ, *Concrete Problems in AI Safety*, ArXiv e-prints, (2016).

[18] S. T. ANDERSON AND R. G. NEWELL, *Information programs for technology adoption: the case of energy-efficiency audits*, Resource and Energy economics, 26 (2004), pp. 27–50.

[19] L. ANDRÉS, D. CUBERES, M. DIOUF, AND T. SEREBRISKY, *The diffusion of the Internet: A cross-country analysis*, Telecommunications policy, 34 (2010), pp. 323–340.

[20] D. ANDREWS, C. CRISCUOLO, AND P. N. GAL, *Frontier firms, technology diffusion and public policy: Micro evidence from OECD countries*, tech. rep., OECD, 2015.

[21] D. ANDREWS, G. NICOLETTI, AND C. TIMILIOTIS, *Digital technology diffusion: A matter of capabilities, incentives or both?*, tech. rep., OECD, July 2018.

[22] M. ARNTZ, T. GREGORY, AND U. ZIERAHN, *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis*, Tech. Rep. 189, OECD Social, Employment, and Migration Working Papers, Paris, France, May 2016.

[23] AUSTRALIA. DEPARTMENT OF EMPLOYMENT, SKILLS, SMALL AND FAMILY BUSINESS. FUTURE OF WORK TASKFORCE., *Reskilling Australia: A data-driven approach*, tech. rep., Australian Government, 2019.

[24] AUSTRALIAN BUREAU OF STATISTICS, *6107.0 - Information paper: Outcomes of the Labour Household Surveys Content Review, 2012*, tech. rep., Commonwealth of Australia, July 2012.

[25] ——, *ANZSCO – Australian and New Zealand Standard Classification of Occupations, 2013, Version 1.2 - 1220.0*.
https://www.abs.gov.au/ausstats/abs@.nsf/0/E3031B89999B4582CA2575DF002DA702?opendocument#:~:text=The%20structure%20of%20ANZSCO%20has,grouped%20into%20'minor%20groups'., June 2013.
Accessed: 2019-8-1.

[26] ——, *Retrenchment Data from Labour Force Survey*.
https://www.abs.gov.au/AUSSTATS/abs@.nsf/Previousproducts/6291.0.55.003Main%20Features9Nov%202015, Dec. 2015.
Accessed: 2020-11-28.

[27] ——, *Business Use of Information Technology*.
https://www.abs.gov.au/statistics/industry/technology-and-innovation/business-use-information-technology/latest-release, July 2017.
Accessed: 2020-9-21.

[28] ——, *Average Weekly Earnings, Australia, Nov 2018 - 6302.0*.
https://www.abs.gov.au/ausstats/abs@.nsf/mf/6302.0, 2018.
Accessed: 2019-8-1.

[29] ——, *Underemployment in Australia: 6202.0 - Labour Force, Australia, September 2018*.
https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/6202.0main+features10September%202018, Oct. 2018.
Accessed: 2019-8-13.

[30] ——, *Employee Earnings and Hours, Australia, May 2018 - 6306.0*, Jan. 2019.

[31] ——, *Labour Force, Australia, Detailed, Quarterly, May 2019 - 6291.0.55.003*, June 2019.

[32] ——, *UNIT GROUP 2124 JOURNALISTS AND OTHER WRITERS*.
https://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/F4496F62EA226B67CA2584A8000E79B9?opendocument, Nov. 2019.
Accessed: 2020-10-20.

[33] ——, *Labour Force, Australia*.
https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release, Nov. 2020.
Accessed: 2020-11-23.

[34] ——, *Labour Force, Australia*.
https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release, Sept. 2020.
Accessed: 2020-10-8.

[35] ——, *Labour Force, Australia, Detailed, Quarterly*.
https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia-detailed-quarterly/latest-release, Mar. 2020.
Accessed: 2020-11-30.

[36] AUSTRALIAN FEDERAL DEPARTMENT OF EDUCATION AND TRAINING, *ucube: Higher education statistics*, 2018.
Title of the publication associated with this dataset: Completions.

[37] AUSTRALIAN FEDERAL DEPARTMENT OF EDUCATION, SKILLS AND EMPLOYMENT, *ANZSCO to O*NET concordance*.

[38] AUSTRALIAN GOVERNMENT, *Skill Shortage Research Methodology*.
https://docs.employment.gov.au/documents/skill-shortage-research-methodology-0, Nov. 2018.
Accessed: 2020-2-9.

[39] AUSTRALIAN GOVERNMENT, DEPARTMENT EDUCATION SKILLS AND EMPLOYMENT, *A statistical snapshot of women in the australian workforce*.

https://www.employment.gov.au/newsroom/statistical-snapshot-women-australian-workforce.
Accessed: 2020-11-25.

[40] AUSTRALIAN GOVERNMENT DEPARTMENT OF HUMAN SERVICES, *Newstart Allowance Income and asset limits*.
https://www.humanservices.gov.au/individuals/enablers/newstart-allowance-income-and-asset-limits/36816.
Accessed: 2018-9-9.

[41] D. AUTOR AND B. PRICE, *The changing task composition of the US labor market: An update of Autor, Levy, and Murnane (2003)*, MIT Paper, 21 (2013).

[42] D. H. AUTOR, *Why Are There Still So Many Jobs? The History and Future of Workplace Automation*, Journal of Economic Perspectives, 29 (2015), pp. 3–30.

[43] D. H. AUTOR, D. DORN, AND G. H. HANSON, *The China Syndrome: Local Labor Market Effects of Import Competition in the United States*, Am. Econ. Rev., 103 (2013), pp. 2121–2168.

[44] D. H. AUTOR AND M. J. HANDEL, *Putting tasks to the test: Human capital, job tasks, and wages*, Journal of labor Economics, 31 (2013), pp. S59–S96.

[45] D. H. AUTOR, L. F. KATZ, AND A. B. KRUEGER, *Computing inequality: have computers changed the labor market?*, The Quarterly journal of economics, 113 (1998), pp. 1169–1213.

[46] D. H. AUTOR, F. LEVY, AND R. J. MURNANE, *Skill Demand, Inequality, and Computerization: Connecting the Dots*, in Technology, Growth, and the Labor Market, D. K. Ginther, M. Zavodny, and L. H. Foley, eds., Springer US, Boston, MA, 2003, pp. 107–129.

[47] ——, *The skill content of recent technological change: An empirical exploration*, The Quarterly Journal of Economics, 118 (2003), pp. 1279–1133.

[48] M. N. BAILY AND B. P. BOSWORTH, *US Manufacturing: Understanding Its Past and Its Potential Future*, Journal of Economic Perspectives., 28 (2014), pp. 3–26.

[49] H. BAKHSHI, J. M. DOWNING, M. A. OSBORNE, AND P. SCHNEIDER, *The Future of Skills: Employment in 2030*, tech. rep., Nesta, 2017.

[50] P. BARNES AND S. KENNARD, *Skill and Australia's Productivity Surge*, tech. rep., Productivity Commission, 2002.

[51] B. S. BARNOW AND J. SMITH, *Employment and Training Programs*, in Economics of Means-Tested Transfer Programs in the United States, R. A. Moffitt, ed., vol. 2, University of Chicago Press, Nov. 2016, pp. 127–234.

[52] A. BASSANINI AND A. GARNERO, *Dismissal protection and worker flows in OECD countries: Evidence from cross-country / cross-industry data*, Labour Economics, 21 (2013), pp. 25–41.

[53] W. J. BAUMOL, S. A. B. BLACKMAN, AND E. N. WOLFF, *Productivity and American Leadership: The Long View*, The MIT Press, first edition ed., Nov. 1989.

[54] T. BAYES AND R. PRICE, *An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S*, Philosophical Transactions (1683-1775), 53 (1763), pp. 370–418.

[55] BEA, *National Income and Product Accounts: 1950-2010*.
https://apps.bea.gov/iTable/index_nipa.cfm, 2012.
Accessed: 2018-10-24.

[56] P. BEAUDRY, M. DOMS, AND E. LEWIS, *Endogenous skill bias in technology adoption: City-level evidence from the IT revolution*, tech. rep., National Bureau of Economic Research, 2006.

[57] N. BECHICHII, R. GRUNDKEI, S. JAMETI, AND M. SQUICCIARINI, *Moving between jobs: an analysis of occupation distances and skill needs*, Tech. Rep. 52, OECD, June 2018.

[58] G. BECKER, *Human capital*, Columbia University: Columbia University Press, 1964.

[59] G. S. BECKER, K. M. MURPHY, AND R. TAMURA, *Human Capital, Fertility, and Economic Growth*, Journal of Political Economy, 98 (1990), pp. S12–S37.

[60] W. BEDINGFIELD, *Coronavirus news fatigue is real and it could become a big problem*.
https://www.wired.co.uk/article/coronavirus-news-fatigue, Apr. 2020.
Accessed: 2020-4-22.

[61] A. BEER, *Structural adjustment programmes and regional development in Australia*, Local Economy, 30 (2015), pp. 21–40.

[62] L. BELLMANN AND O. HÜBLER, *The skill shortage in German establishments before, during and after the great recession*, Jahrbücher für Nationalökonomie und Statistik, 234 (2014), pp. 800–828.

[63] J. BENNETT AND S. MCGUINNESS, *Assessing the impact of skill shortages on the productivity performance of high-tech firms in Northern Ireland*, Applied Economics, 41 (2009), pp. 727–737.

[64] J. BENTON, *On a rough day for American newspapers, investors aren't buying gannett's story and tribune's not done chopping*, (2020).

[65] J. BERGSTRA AND Y. BENGIO, *Random search for hyper-parameter optimization*, Journal of machine learning research, 13 (2012), pp. 281–305.

[66] E. BERMAN, J. BOUND, AND Z. GRILICHES, *Changes in the demand for skilled labor within US manufacturing: evidence from the annual survey of manufactures*, The Quarterly Journal of Economics, 109 (1994), pp. 367–397.

[67] J. BESSEN, *Technology adoption costs and productivity growth: The transition to information technology*, Review of Economic Dynamics, (2002).

[68] J. BESSEN, *Learning by Doing: The Real Connection between Innovation, Wages, and Wealth*, Yale University Press, Apr. 2015.

[69] J. E. BESSEN, S. M. IMPINK, R. SEAMANS, AND L. REICHENSPERGER, *The Business of AI Startups*.
Nov. 2018.

[70] A. BLAKE, *Dynamics of data science skills*, tech. rep., The Royal Society, May 2019.

[71] A. S. BLINDER AND A. B. KRUEGER, *Alternative measures of offshorability: a survey approach*, Journal of Labor Economics, 31 (2013), pp. S97–S128.

[72] D. BÖGENHOLD AND U. FACHINGER, *Blurred boundaries of journalism: multiple employment in the media industry and the hybridity of occupational work*, International Journal of Arts and Commerce, 2 (2013), pp. 171–183.

[73] G. BOOLE, *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*, Macmillan, NY, New York, 1854.

[74] G. J. BORJAS AND J. C. VAN OURS, *Labor economics*, McGraw-Hill/Irwin Boston, 2010.

[75] J. BORLAND, *Microeconomic Reform and Displaced Workers: An Introduction*, Tech. Rep. 385, 1998.

[76] M. BOSELEY AND C. KNAUS, *Australia's coronavirus social distancing rules explained: state by state guidelines*, The Guardian, (2020).

[77] N. BOSTROM, *Superintelligence: Paths, Dangers, Strategies*, OUP Oxford, July 2014.

[78] P. BRANCO, L. TORGO, AND R. RIBEIRO, *A survey of predictive modelling under imbalanced distributions*, (2015).

[79] T. F. BRESNAHAN, E. BRYNJOLFSSON, AND L. M. HITT, *Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence*, The Quarterly Journal of Economics, 117 (2002), pp. 339–376.

[80] T. F. BRESNAHAN AND M. TRAJTENBERG, *General purpose technologies 'Engines of growth'?*, Journal of Econometrics, 65 (1995), pp. 83–108.

[81] P. J. BROCKWELL, R. A. DAVIS, AND M. V. CALDER, *Introduction to time series and forecasting*, vol. 2, Springer, 2002.

[82] T. B. BROWN, B. MANN, N. RYDER, M. SUBBIAH, J. KAPLAN, P. DHARIWAL, A. NEELAKANTAN, P. SHYAM, G. SASTRY, A. ASKELL, S. AGARWAL, A. HERBERT-VOSS, G. KRUEGER, T. HENIGHAN, R. CHILD, A. RAMESH, D. M. ZIEGLER, J. WU, C. WINTER, C. HESSE, M. CHEN, E. SIGLER, M. LITWIN, S. GRAY, B. CHESS, J. CLARK, C. BERNER, S. MCCANDLISH, A. RADFORD, I. SUTSKEVER, AND D. AMODEI, *Language Models are Few-Shot Learner*, ArXiv e-prints, (2020).

[83] BROWNE, IRENE AND MISRA, JOYA, *The intersection of gender and race in the labor market*, Annual Review of Sociology, 29 (2003), pp. 487–513.

[84] M. BRUCKNER, M. LAFLEUR, AND I. PITTERLE, *The impact of the technological revolution on labour markets and income distribution*, tech. rep., United Nations, July 2017.

[85] G. BRUNELLO AND P. WRUUCK, *Skill shortages and skill mismatch in Europe: A review of the literature*, (2019).

[86] E. BRYNJOLFSSON AND A. MCAFEE, *Why Workers Are Losing the War Against Machines*, The Atlantic, (2011).

[87] E. BRYNJOLFSSON AND A. MCAFEE, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, WW Norton & Company, 2014.

[88] E. BRYNJOLFSSON AND P. MILGROM, *Complementarity in organizations*, The Handbook of Organizational Economics, (2013), pp. 11–55.

[89] E. BRYNJOLFSSON, D. ROCK, AND C. SYVERSON, *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*, in The Economics of Artificial Intelligence: An Agenda, University of Chicago Press, Jan. 2018.

[90] ——, *The Productivity J-Curve: How Intangibles Complement General Purpose Technologies*, Tech. Rep. w25148, National Bureau of Economic Research, Oct. 2018.

[91] J. N. BUDARICK AND G.-S. HAN, *Minorities and Media:: Producers, Industries, Audiences*, Palgrave Macmillan, 2017.

[92] J. BUGHIN, E. HAZAN, S. LUND, P. DAHLSTRÖM, A. WIESINGER, AND A. SUBRAMANIAM, *Skill shift: Automation and the future of the workforce*, tech. rep., McKinsey Global Institute, 2018.

[93] J. BUGHIN, J. SEONG, J. MANYIKA, M. CHUI, AND R. JOSHI, *Notes from the AI frontier: Modeling the impact of AI on the world economy*, tech. rep., McKinsey Global Institute, 2018.

[94] R. CALO, *Artificial Intelligence policy: a primer and roadmap*, UCDL Rev., 51 (2017), p. 399.

[95] P. H. CAPPELLI, *Skill Gaps, Skill Shortages, and Skill Mismatches: Evidence and Arguments for the United States*, ILR Review, 68 (2015), pp. 251–290.

[96] D. CARD, J. KLUVE, AND A. WEBER, *What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations*, Journal of the European Economic Association, 16 (2018), pp. 894–931.

[97] J. W. CAREY, *The communications revolution and the professional communicator*, The Sociological Review, 13 (1965), pp. 23–38.

[98] M. CARLSON, *Establishing the boundaries of journalism‚Äôs public mandate*, C. Peters and M, (2016).

[99] M. CARLSON AND S. C. LEWIS, *Boundaries of journalism: Professionalism, practices and participation*, Routledge, 2015.

[100] CARLSSON, MAGNUS AND ERIKSSON, STEFAN, *Age discrimination in hiring decisions: Evidence from a field experiment in the labor market*, Labour Economics, 59 (2019), pp. 173–183.

[101] A. CARNEVALE, T. JAYASUNDERA, AND D. REPNIKOV, *Understanding Online Job Ads Data*, tech. rep., Georgetown University, 2014.

[102] A. CASERO-RIPOLLÉS, *Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak*, El profesional de la información, 29 (2020), p. e290223.

[103] CEDA, *We need to build skills, not walls: Telstra CEO*.
https://www.ceda.com.au/News-and-analysis/CEDA-Events/We-need-to-build-skills-not-walls-Telstra-CEO, 2019.
Accessed: 2020-2-6.

[104] CEDEFOP, *Shortages and gaps in European enterprises: striking a balance between vocational education and training and the labour market*, CEDEFOP reference series, Luxembourg, 102 (2015).

[105] CEDEFOP, *Insights into skill shortages and skill mismatch: Learning from CEDEFOP's European skills and jobs survey*, 2018.

[106] S. CHALMERS, *9,500 jobs to save one: How Telstra is slashing its workforce to meet NBN challenge*, ABC News, (2019).

[107] T. CHEN, *Introduction to boosted trees*, University of Washington Computer Science, 22 (2014), p. 115.

[108] T. CHEN AND C. GUESTRIN, *XGBoost: A Scalable Tree Boosting System*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, New York, NY, USA, Aug. 2016, Association for Computing Machinery, pp. 785–794.

[109] I. M. COCKBURN, R. HENDERSON, AND S. STERN, *The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis*, in The Economics of Artificial Intelligence: An Agenda, University of Chicago Press, Jan. 2018.

[110] M. COELLI AND J. BORLAND, *Job Polarisation and Earnings Inequality in Australia*, Economic Record, 92 (2016), pp. 1–27.

[111] E. CONNOLLY AND C. LEWIS, *Structural Change in the Australian Economy*, Tech. Rep. September Quarter, Reserve Bank of Australia, 2010.

[112] K. CRAWFORD AND M. WHITTAKER, *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*, Tech. Rep. 1, AI Now, Sept. 2016.

[113] S. CRERAR, *Coronavirus and the crisis in regional news*.
https://www.uts.edu.au/research-and-teaching/our-research/centre-media-transition/news/coronavirus-and-crisis-regional-news, Apr. 2020.
Accessed: 2020-4-27.

[114] T. CROSS, *After Moore's law*, The Economist, (2016).

[115] D. CROWE, *'A level playing field': digital giants will have to pay for news*.
https://www.smh.com.au/politics/federal/a-level-playing-field-digital-giants-will-have-to-pay-for-news-20200419-p54l7q.html, Apr. 2020.
Accessed: 2020-4-20.

[116] CSIRO, *Our Top 10 Inventions*.
https://www.csiro.au/en/About/History-achievements/Top-10-inventions, Mar. 2015.
Accessed: 2018-8-14.

[117] N. DAVIS, *The future relationship between technology and inequality*, tech. rep., Committee for Economic Development of Australia, 2018.

[118] N. DAWSON, S. MOLITORISZ, M.-A. RIZOIU, AND P. FRAY, *Layoffs, inequity and COVID-19: A longitudinal study of the journalism jobs crisis in australia from 2012 to 2020*, Journalism, (2021), p. 1464884921996286.

[119] N. DAWSON AND M.-A. RIZOIU, *Coronavirus infecting Australian jobs: vacancy rates down since early February*, The Conversation, (2020).

[120] N. DAWSON, M.-A. RIZOIU, B. JOHNSTON, AND M.-A. WILLIAMS, *Adaptively selecting occupations to detect skill shortages from online job ads*, in 2019 IEEE International Conference on Big Data (Big Data), IEEE, 2019, pp. 1637–1643.

[121] ——, *Predicting Skill Shortages in Labor Markets: A Machine Learning Approach*, in The 4th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in BigData (IEEE HMData 2021), Apr. 2020.

[122] J. G. DE GOOIJER AND R. J. HYNDMAN, *25 years of time series forecasting*, International journal of forecasting, 22 (2006), pp. 443–473.

[123] DELOITTE ACCESS ECONOMICS, *ACS Australia's Digital Pulse 2018: Driving Australia's international ICT competitiveness and digital growth*, tech. rep., Australian Computer Society, 2018.

[124] DEPARTMENT OF EDUCATION, SKILLS AND EMPLOYMENT, AUSTRALIAN GOVERNMENT, *Historical list of skill shortages in Australia*, 2019.

[125] DEPARTMENT OF EMPLOYMENT, SKILLS, SMALL, FAMILY BUSINESS, AND AUSTRALIAN GOVERNMENT, *Skill shortages*.
https://www.employment.gov.au/skillshortages.
Accessed: 2019-11-1.

[126] DEPARTMENT OF EMPLOYMENT, SKILLS, SMALL AND FAMILY BUSINESS, *Sixty per cent of job vacancies in Australia are advertised online*.
https://www.employment.gov.au/newsroom/sixty-cent-job-vacancies-australia-are-advertised-online, Aug. 2018.
Accessed: 2019-7-7.

[127] DEPARTMENT OF JOBS AND SMALL BUSINESS, *The next generation of employment services - Appendices*, tech. rep., Australian Government, 2018.

[128] DEPARTMENT OF JOBS AND SMALL BUSINESS, COMMONWEALTH OF AUSTRALIA, *Jobactive*.
https://jobactive.gov.au/, 2018.
Accessed: 2018-9-10.

[129] DEPARTMENT OF SERVICES, AUSTRALIAN GOVERNMENT, *Income and assets tests*.
https://www.servicesaustralia.gov.au/individuals/services/centrelink/jobseeker-payment/how-much-you-can-get/income-and-assets-tests, 2020.
Accessed: 2020-11-30.

[130] DEPARTMENT OF SOCIAL SERVICES AND MELBOURNE INSTITUTE OF APPLIED ECONOMIC AND SOCIAL RESEARCH, *The Household, Income and Labour Dynamics in Australia (HILDA) Survey, RESTRICTED RELEASE 18 (Waves 1-18)*, 2020.

[131] V. DHAR, *Data science and prediction*, Commun. ACM, 56 (2013), pp. 64–73.

[132] Z. DHILLON AND N. CASSIDY, *Labour Market Outcomes for Younger People: Bulletin – June Quarter 2018*, tech. rep., Reserve Bank of Australia, 2018.

[133] K. DOCTOR, *Newsonomics: What was once unthinkable is quickly becoming reality in the destruction of local news*.
https://www.niemanlab.org/2020/03/newsonomics-what-was-once-unthinkable-is-quickly-becoming-reality-in-the-destruction-of-local-news/, Mar. 2020.
Accessed: 2020-4-19.

[134] L. DORMEHL, *What is an artificial neural network? Here's everything you need to know*.
https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/, Sept. 2018.
Accessed: 2018-10-24.

[135] C. DOUGHERTY, *Prometheus*, Taylor & Francis, 2006.

[136] H. DURRANT-WHYTE, L. MCCALMAN, S. O'CALLAGHAN, A. REID, AND D. STEINBERG, *The impact of computerisation and automation on future employment*, tech. rep., CEDA, 2015.

[137] R. EDMONDS, *A fresh publishing study argues that paid digital subscriptions are the only avenue to growth*.

https://www.poynter.org/business-work/2020/a-fresh-publishing-study-argues-that-paid-digital-subscriptions-are-the-only-avenue-to-growth/, Sept. 2020. Accessed: 2020-10-24.

[138] L. EINAV AND J. LEVIN, *Economics in the age of big data*, Science, 346 (2014), p. 1243089.

[139] FAETHM, *Australian essential and non-essential occupations during COVID-19 2020*.
https://github.com/Faethm-ai/open-data/blob/main/essential-occupations-AUS/ANZSCO_4digit_essential_v_nonessential.csv, 2021.
Accessed: 2021-5-7.

[140] P. FAJNZYLBER, D. LEDERMAN, AND N. LOAYZA, *Inequality and Violent Crime*, The Journal of Law and Economics, 45 (2002), pp. 1–39.

[141] B. FALLICK, *A Review of the Recent Empirical Literature on Displaced Workers*, ILR Review, 50 (1996), pp. 5–16.

[142] N. FENTON, *Deregulation or democracy? New media, news, neoliberalism and the public interest*, Continuum, 25 (2011), pp. 63–72.

[143] A. FERNÁNDEZ, S. GARCÍA, M. GALAR, R. C. PRATI, B. KRAWCZYK, AND F. HERRERA, *Learning from Imbalanced Data Sets*, Springer, 2018.

[144] C. FISHER, *'Watchdog' to 'spin-doctor': Challenges and benefits*, Australian Journalism Review, 36 (2014), p. 145.

[145] C. FISHER, S. PARK, G. LEE, J.AND FULLER, AND Y. SANG, *Digital news report: Australia 2019*, tech. rep., News and Media Research Centre, Mar. 2019.

[146] FOLEY, MERAIAH AND WILLIAMSON, SUE, *Does anonymising job applications reduce gender bias? understanding managers,Äô perspectives*, Gender in Management: An International Journal, (2018).

[147] J. FORTH AND G. MASON, *Do ICT skill shortages hamper firms,Äô performance? Evidence from UK benchmarking surveys*, National Institute of Economic and Social Research, Discussion Paper, 281 (2006).

[148] M. R. FRANK, D. AUTOR, J. E. BESSEN, E. BRYNJOLFSSON, M. CEBRIAN, D. J. DEMING, M. FELDMAN, M. GROH, J. LOBO, E. MORO, D. WANG, H. YOUN,

AND I. RAHWAN, *Toward understanding the impact of artificial intelligence on labor*, Proceedings of the National Academy of Sciences of the United States of America, (2019), p. 201900949.

[149] R. B. FREEMAN, *Overinvestment in college training?*, Journal of human resources, (1975), pp. 287–311.

[150] FREEMAN, CHRISTOPHER AND PEREZ, CARLOTA, *Structural crises of adjustment: business cycles*, in Technical Change and Economic Theory, Giovanni Dosi, Christopher Freeman, Richard Nelson, Gerald Silverberg, and Luc Soete, ed., Londres: Pinter, Italy: Sant'Anna School of Advanced Studies, 1988.

[151] C. B. FREY AND M. A. OSBORNE, *The future of employment: How susceptible are jobs to computerisation?*, Technological Forecasting and Social Change, 114 (2017), pp. 254–280.

[152] M. L. FROGNER, *Skills shortages: An examination of the supply and demand for skills, and the links between skills shortages and the labour market and earnings*, Labour Market Trends, 110 (2002), pp. 17–28.

[153] J. FURMAN, *Should We Be Reassured If Automation in the Future Looks Like Automation in the Past?*, in The Economics of Artificial Intelligence: An Agenda, A. K. Agrawal, J. Gans, and A. Goldfarb, eds., University of Chicago Press, 2018.

[154] J. FURMAN, J. HOLDREN, C. MUÑOZ, M. SMITH, AND J. ZIENTS, *Artificial Intelligence, Automation, and the Economy*, tech. rep., Executive Office of the President, Dec. 2016.

[155] J. FURMAN AND R. SEAMANS, *AI and the Economy*, in Innovation Policy and the Economy, Volume 19, University of Chicago Press, May 2018.

[156] R. E. GALLMAN AND T. J. WEISS, *The Service Industries in the Nineteenth Century*, in Production and Productivity in the Service Industries, V. R. Fuchs, ed., NBER, 1969, pp. 287–381.

[157] A. GARDINER, C. AASHEIM, P. RUTNER, AND S. WILLIAMS, *Skill Requirements in Big Data: A Content Analysis of Job Advertisements*, Journal of Computer Information Systems, 58 (2018), pp. 374–384.

[158] C. GATHMANN AND U. SCHÖNBERG, *How general is human capital? a task-based approach*, Journal of Labor Economics, 28 (2010), pp. 1–49.

[159] F. W. GEELS, *The dynamics of transitions in socio-technical systems: A multi-level analysis of the transition pathway from horse-drawn carriages to automobiles (1860–1930)*, Technology Analysis & Strategic Management, 17 (2005), pp. 445–476.

[160] A. GIABELLI, L. MALANDRI, F. MERCORIO, M. MEZZANZANICA, AND A. SEVESO, *NEO: A Tool for Taxonomy Enrichment with New Emerging Occupations*, in The Semantic Web – ISWC 2020, Springer International Publishing, 2020, pp. 568–584.

[161] G. GILFILLAN, *Characteristics and use of casual employees in Australia*, Tech. Rep. Research paper series, 2017–18, Parliament of Australia, Jan. 2018.

[162] B. GOGGIN, *7,800 people have lost their jobs so far this year in a media landslide*.
https://www.businessinsider.com.au/2019-media-layoffs-job-cuts-at-buzzfeed-huffpost-vice-details-2019-2, Feb. 2019.
Accessed: 2020-3-10.

[163] C. D. GOLDIN, *Human capital*, (2016).

[164] C. D. GOLDIN AND L. F. KATZ, *The race between education and technology*, Harvard University Press, 2009.

[165] I. GOODFELLOW, Y. BENGIO, A. COURVILLE, AND Y. BENGIO, *Deep learning*, vol. 1, MIT press Cambridge, 2016.

[166] GOOGLE, *Google cloud AutoML*.
https://cloud.google.com/automl.
Accessed: 2020-9-21.

[167] M. GOOS, A. MANNING, AND A. SALOMONS, *Explaining job polarization: Routine-biased technological change and offshoring*, American economic review, 104 (2014), pp. 2509–26.

[168] A. GROMOV, A. MASLENNIKOV, N. DAWSON, K. MUSIAL, AND K. KITTO, *Curriculum profile: modelling the gaps between curriculum and the job market*, in Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020), 2020, pp. 610–614.

[169] R. GRUNDKE, S. JAMET, M. KALAMOVA, AND M. SQUICCIARINI, *Having the right mix: The role of skill bundles for comparative advantage and industry performance in GVCs*, (2017).

[170] L. GUO AND Y. VOLZ, *(re) defining journalistic expertise in the digital transformation: A content analysis of job announcements*, Journalism Practice, 13 (2019), pp. 1294–1315.

[171] B. H. HALL AND B. KHAN, *Adoption of new technology*, tech. rep., National Bureau of Economic Research, 2003.

[172] T. HANITZSCH, F. HANUSCH, J. RAMAPRASAD, AND A. S. DE BEER, *Worlds of journalism: Journalistic cultures around the globe*, Columbia University Press, 2019.

[173] Y. N. HARARI, *Homo Deus: A Brief History of Tomorrow*, VINTAGE ARROW - MASS MARKET, Apr. 2017.

[174] HARRIS, ROB AND BAGSHAW, ERYK, *Strict new controls announced as morrison government tries to limit spread of COVID-19*, The Sydney Morning Herald, (2020).

[175] J. HASKEL AND C. MARTIN, *Do skill shortages reduce productivity? Theory and evidence from the United Kingdom*, The Economic Journal, 103 (1993), pp. 386–394.

[176] ——, *Skill shortages, productivity growth and wage inflation*, Acquiring Skills: Market Failures: Their Symptoms and Policy Responses, (1996), pp. 147–174.

[177] D. HASSABIS AND D. SILVER, *AlphaGo Zero: Learning from scratch*. https://deepmind.com/blog/alphago-zero-learning-scratch/, Oct. 2017. Accessed: 2017-10-28.

[178] J. HASSLER, J. V. RODRIGUEZ MORA, K. STORESLETTEN, AND F. ZILIBOTTI, *A positive theory of geographic mobility and social insurance*, International Economic Review, 46 (2005), pp. 263–303.

[179] J. HAYES AND L. RUBBO, *$50m coronavirus bailout announced for regional media*, ABC News, (2020).

[180] J. HEALY, K. MAVROMARAS, AND P. J. SLOANE, *Adjusting to skill shortages in Australian SMEs*, Applied Economics, 47 (2015), pp. 2470–2487.

[181] A. HEATH, *Structural Change in Australian Industry: The Role of Business Services*.
Economic Society of Australia, Sept. 2017.

[182] ——, *Skills, Technology and the Future of Work*.
Career Education Association of Victoria and Victorian Commercial Teachers Association Work Futures Conference, Mar. 2020.

[183] E. HELPMAN AND A. RANGEL, *Adjusting to a New Technology: Experience and Training*, J. Econ. Growth, 4 (1999), pp. 359–383.

[184] T. HEY, S. TANSLEY, AND K. TOLLE, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 1 edition ed., Oct. 2009.

[185] C. A. HIDALGO, B. KLINGER, A.-L. BARABÁSI, AND R. HAUSMANN, *The product space conditions the development of nations*, Science, 317 (2007), pp. 482–487.

[186] P. L. HIGGS AND S. D. CUNNINGHAM, *Australia's creative economy: Mapping methodologies*, (2007).

[187] IBM, *10 Key Marketing Trends for 2017: IBM and Ideas for Exceeding Customer Expectations*, tech. rep., IBM, Dec. 2016.

[188] I. IM, S. HONG, AND M. S. KANG, *An international comparison of technology adoption: Testing the UTAUT model*, Information & management, 48 (2011), pp. 1–8.

[189] B. F. INGRAM AND G. R. NEUMANN, *The returns to skill*, Labour economics, 13 (2006), pp. 35–59.

[190] E. IZADI, *With cuts at Vice, Quartz and BuzzFeed, even media's savviest digital players are hurting*, 2020.
Accessed: 2020-10-22.

[191] C. JI, X. ZOU, Y. HU, S. LIU, L. LYU, AND X. ZHENG, *XG-SF: An XGBoost classifier based on shapelet features for time series classification*, Procedia Comput. Sci., 147 (2019), pp. 24–28.

[192] M. I. JORDAN AND T. M. MITCHELL, *Machine learning: Trends, perspectives, and prospects*, Science, 349 (2015), pp. 255–260.

[193] B. JOSEPHI AND M. OLLER ALONSO, *Re-examining age: Journalism's reliance on the young*, Journalism, (2018), p. 1464884918800077.

[194] B. JOVANOVIC AND P. L. ROUSSEAU, *General purpose technologies*, in Handbook of economic growth, vol. 1, Elsevier, 2005, pp. 1181–1224.

[195] T. A. JUDGE AND R. D. BRETZ, *Effects of work values on job choice decisions*, Journal of Applied Psychology, 77 (1992), pp. 261–271.

[196] P. JUNANKAR, *Was there a skills shortage in Australia?*, (2009).

[197] A. KALOGEROPOULOS, R. FLETCHER, AND R. K. NIELSEN, *Initial surge in news use around coronavirus in the UK has been followed by significant increase in news avoidance*, Available at SSRN 3609184, (2020).

[198] E. KARAHANNA, D. W. STRAUB, AND N. L. CHERVANY, *Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs*, MIS quarterly, (1999), pp. 183–213.

[199] L. F. KATZ ET AL., *Changes in the wage structure and earnings inequality*, in Handbook of labor economics, vol. 3, Elsevier, 1999, pp. 1463–1555.

[200] H. KENDIG, P. MCDONALD, AND J. PIGGOTT, *Population ageing and Australia's future*, ANU Press, 2016.

[201] T. KENNEDY, M. RAE, A. SHERIDAN, AND A. VALADKHANI, *Reducing gender wage inequality increases economic prosperity for all: Insights from Australia*, Econ. Anal. Policy, 55 (2017), pp. 14–24.

[202] M. L. KERN, P. X. MCCARTHY, D. CHAKRABARTY, AND M.-A. RIZOIU, *Social media-predicted personality traits and values can help match people to their ideal jobs*, Proceedings of the National Academy of Sciences of the United States of America, 116 (2019), pp. 26459–26464.

[203] J. M. KEYNES, *Economic Possibilities for our Grandchildren*, in Essays in Persuasion, Harcourt Brace, New York, 1932, pp. 358–373.

[204] E. KOEHN, *'We're just getting started': Amazon Australia revenue surges to $292m*, The Sydney Morning Herald, (2019).

[205] A. KORINEK AND D. X. NG, *The Macroeconomics of Superstars*. Nov. 2017.

[206] A. KORINEK AND J. E. STIGLITZ, *Artificial Intelligence and its Implications for Income Distribution and Unemployment*, NBER WORKING PAPER SERIES, (2017), p. 44.

[207] S. KUZNETS, *Modern Economic Growth: Findings and Reflections*, American Economic Review, 63 (1973), pp. 247–258.

[208] K. F. KYLE DEMARIA AND K. WARDRIP, *Exploring a skills-based approach to occupational mobility*, tech. rep., Federal Reserve Banks of Philadelphia, and Cleveland, 2020.

[209] N. LAB, *About Nieman Lab*. https://www.niemanlab.org/about/, 2020. Accessed: 2020-3-10.

[210] P. R. G. LAYARD, R. LAYARD, S. J. NICKELL, AND R. JACKMAN, *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford University Press, 2005.

[211] C. K. LEE, *Analysis of skill requirements for systems analysts in Fortune 500 organizations*, Journal of Computer Information Systems, 45 (2005), pp. 84–92.

[212] Y. LI, *Deep Reinforcement Learning: An Overview*, (2017).

[213] LINKEDIN ECONOMIC GRAPH TEAM, *LinkedIn Workforce Report - United States*, tech. rep., LinkedIn, 2018.

[214] W. LOOSEN, *The notion of the ‚Äúblurring boundaries‚Äù journalism as a (de-) differentiated phenomenon*, Digital Journalism, 3 (2015), pp. 68–84.

[215] P. MAARES AND F. HANUSCH, *Exploring the boundaries of journalism: Instagram micro-bloggers in the twilight zone of lifestyle journalism*, Journalism, 21 (2020), pp. 262–278.

[216] J. MACNAMARA, *Journalism and PR: Unpacking 'Spin', Stereotypes and Media Myths*, Peter Lang Publishing, 2014.

[217] ——, *The continuing convergence of journalism and PR: New insights for ethical practice from a three-country study of senior practitioners*, Journalism & Mass Communication Quarterly, 93 (2016), pp. 118–141.

[218] R. MADDOCK AND I. W. MCLEAN, eds., *The Australian Economy in the Long Run*, Cambridge University Press, Aug. 1987.

[219] J. B. MADSEN, *Australian economic growth and its drivers since European settlement*, in The Cambridge economic history of Australia, Cambridge University Press, 2015, pp. 29–51.

[220] S. MAKRIDAKIS, *Accuracy measures: theoretical and practical concerns*, International Journal of Forecasting, 9 (1993), pp. 527–529.

[221] ——, *The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms*, Futures, 90 (2017), pp. 46–60.

[222] J. W. MAMER AND K. F. MCCARDLE, *Uncertainty, Competition, and the Adoption of New Technology*, Management Science, 33 (1987), pp. 161–177.

[223] G. L. MANGUM AND H. R. BOWEN, *Technology and the American Economy*, tech. rep., National Commission on Technology, Automation, and Economic Progress, Washington DC, 1966.

[224] MANPOWERGROUP, *Solving the Talent Shortage: 2018 Talent Shortage Survey*, tech. rep., ManpowerGroup, 2018.

[225] J. MANYIKA, M. CHUI, B. BROWN, J. BUGHIN, R. DOBBS, C. ROXBURGH, AND A. H. BYERS, *Big data: The next frontier for innovation, competition, and productivity*, tech. rep., McKinsey Global Institute, 2011.

[226] J. MANYIKA, S. LUND, M. CHUI, J. BUGHIN, J. WOETZEL, P. BATRA, R. KO, AND S. SANGHVI, *Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation*, tech. rep., McKinsey Global Institute, 2017.

[227] W. MARKOW, S. BRAGANZA, B. TASKA, S. M. MILLER, AND D. HUGHES, *The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market*, tech. rep., Burning Glass Technologies, 2017.

[228] J. McCarthy, *What is Artificial Intelligence?*, tech. rep., Stanford, 2007.

[229] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, AI Magazine, 27 (1955).

[230] S. McGuinness, K. Pouliakas, and P. Redmond, *How useful is the concept of skills mismatch?*, (2017).

[231] A. Meade, *Dozens of Australian newspapers stop printing as coronavirus crisis hits advertising*, The Guardian, (2020).

[232] ——, *News Corp announces end of more than 100 Australian print newspapers in huge shift to digital*, The Guardian, (2020).

[233] ——, *News Corp Australia warns of coronavirus crisis job cuts as smaller regional papers close*, The Guardian, (2020).

[234] B. Milanovic, *Global Inequality: A New Approach for the Age of Globalization*, Harvard University Press, reprint edition ed., Apr. 2018.

[235] J. Mincer, *Human capital, technology, and the wage structure: what do time series show?*, tech. rep., National Bureau of Economic Research, 1991.

[236] D. Mishra and U. Deichmann, *World Development Report 2016: Digital Dividends*, tech. rep., The World Bank, June 2016.

[237] J. Mokyr, C. Vickers, and N. L. Ziebarth, *The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?*, Journal of Economic Perspectives, 29 (2015), pp. 31–50.

[238] K. S. Moorthy, *Using game theory to model competition*, Journal of Marketing Research, 22 (1985), pp. 262–282.

[239] E. Moro, M. R. Frank, A. Pentland, A. Rutherford, M. Cebrian, and I. Rahwan, *Universal resilience patterns in labor markets*, Nature Communications, 12 (2021), pp. 1–8.

[240] V. C. Müller and N. Bostrom, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, in Fundamental Issues of Artificial Intelligence, Synthese Library, Springer, Cham, 2016, pp. 555–572.

[241] G. MURTOUGH AND M. WAITE, *Unemployment and Re-employment of Displaced Workers*, tech. rep., Productivity Commission Staff Research Paper, Oct. 2000.

[242] L. NEDELKOSKA, D. DIODATO, AND F. NEFFKE, *Is our human capital general enough to withstand the current wave of technological change?*, tech. rep., Center for International Development, Harvard University, 2018.

[243] L. NEDELKOSKA AND F. NEFFKE, *Skill Mismatch and Skill Transferability: Review of Concepts and Measurements*, Papers in Evolutionary Economic Geography, (2019).

[244] L. NEDELKOSKA, F. NEFFKE, AND S. WIEDERHOLD, *Skill mismatch and the costs of job displacement*, in Annual Meeting of the American Economic Association, 2015.

[245] S. NICKELL, D. NICOLITSAS, ET AL., *Human capital, investment and innovation: what are the connections?*, tech. rep., 1997.

[246] N. J. NILSSON, *The Quest for Artificial Intelligence*, Cambridge University Press, 1 edition ed., Oct. 2009.

[247] S. NOMURA, S. IMAIZUMI, A. C. AREIAS, AND F. YAMAUCHI, *Toward Labor Market Policy 2.0: The Potential for Using Online Job-Portal Big Data to Inform Labor Market Policies in Indi*, Policy Research Working Papers, The World Bank, Feb. 2017.

[248] L. NORTH, *The gendered newsroom: how journalists experience the changing world of media*, Hampton Press, 2009.

[249] ——, *The gender of "soft" and "hard" news: Female journalists' views on gendered story allocations*, Journalism Studies, 17 (2016), pp. 356–373.

[250] A. NORTON AND I. CHERASTIDTHAM, *Mapping Australian higher education 2018*, tech. rep., Grattan Institute, 2018.

[251] OECD, *Education at a Glance 2003*, tech. rep., OECD, 2003.

[252] ——, *Back to Work: Canada - Improving the Re-employment Prospects of Displaced Workers*, tech. rep., OECD Paris, 2015.

[253] ——, *World Indicators of Skills for Employment (WISE) database*, 2015.

[254] ——, *Back to Work*, OECD Publishing, Paris, Dec. 2016.

[255] ——, *Back to Work: Australia: Improving the re-employment prospects of displaced workers*, OECD Publishing, Paris, 2016.

[256] ——, *Education at a Glance 2017*, tech. rep., Organization for Economic Cooperation and Development, 2017.

[257] ——, *Getting Skills Right: Skills for Jobs Indicators*, OECD Publishing, Paris, 2017.

[258] ——, *The Pursuit of Gender Equality: An Uphill Battle*, OECD publishing, Paris, 2017.

[259] ——, *OECD - Income inequality*.
https://data.oecd.org/inequality/income-inequality.htm, 2018.
Accessed: 2020-8-9.

[260] ——, *OECD - Poverty rate*.
https://data.oecd.org/inequality/poverty-rate.htm, 2018.
Accessed: 2018-10-24.

[261] ——, *OECD Employment Outlook 2018*, OECD Publishing, Paris, 2018.

[262] ——, *Getting Skills Right: Future-Ready Adult Learning Systems*, OECD Publishing, Paris, 2019.

[263] ——, *OECD Skills Strategy 2019 - Skills to Shape a Better Future*, tech. rep., OECD, May 2019.

[264] ——, *Long-term care workforce: caring for the ageing population with dignity*.
https://www.oecd.org/health/health-systems/long-term-care-workforce.htm, 2020.
Accessed: 2020-2-8.

[265] ——, *OECD Employment Outlook 2020*, 2020.

[266] ——, *Unemployment Rates, OECD - Updated: September 2020 - OECD*.
https://www.oecd.org/newsroom/unemployment-rates-oecd-update-september-2020.htm, Oct. 2020.
Accessed: 2020-11-23.

[267] R. K. OLSEN, V. PICKARD, AND O. WESTLUND, *Communal news work: Covid-19 calls for collective funding of journalism*, Digital Journalism, (2020), pp. 1–8.

[268] T. O'REGAN AND C. YOUNG, *Journalism by numbers: trajectories of growth and decline of journalists in the Australian census 1961–2016*, Media International Australia, 172 (2019), pp. 13–32.

[269] L. ORSEAU AND M. RING, *Space-Time Embedded Intelligence*, in Artificial General Intelligence, J. Bach, B. Goertzel, and M. Iklé, eds., Springer, 2012, pp. 209–218.

[270] P. ORZECHOWSKI, W. LA CAVA, AND J. H. MOORE, *Where Are We Now? A Large Benchmark Study of Recent Symbolic Regression Methods*, in Genetic and Evolutionary Computation Conference, GECCO'18, ACM, 2018, pp. 1183–1190.

[271] PAGER, DEVAH AND BONIKOWSKI, BART AND WESTERN, BRUCE, *Discrimination in a low-wage labor market: A field experiment*, American Sociological Review, 74 (2009), pp. 777–799.

[272] S. PARK, C. FISHER, J. Y. LEE, AND K. MCGUINNESS, *Covid-19: Australian news and misinformation*, tech. rep., 2020.

[273] B. M. PAVLYSHENKO, *Linear, machine learning and probabilistic approaches for time series analysis*, in Data Stream Mining Processing (DSMP), Aug. 2016, pp. 377–381.

[274] R. PEARCE, *Telstra seeks to boost tech skills pipeline*. https://www.computerworld.com/article/3462420/telstra-seeks-to-boost-tech-skills-pipeline.html, Oct. 2019. Accessed: 2020-2-6.

[275] G. PERINO AND T. REQUATE, *Does more stringent environmental regulation induce or reduce technology adoption? When the rate of technology adoption is inverted U-shaped*, Journal of Environmental Economics and Management, 64 (2012), pp. 456–467.

[276] R. PERRAULT, Y. SHOHAM, E. BRYNJOLFSSON, J. CLARK, J. ETCHEMENDY, B. GROSZ, T. LYONS, J. MANYIKA, S. MISHRA, AND J. C. NIEBLES, *The AI Index 2019 Annual Report*, tech. rep., Human-Centered AI Institute, Stanford University, Dec. 2019.

[277] A. PERRIN, *Social media usage*, Pew research center, (2015), pp. 52–68.

[278] PIJI, *The Australian Newsroom Mapping project*.
https://anmp.piji.com.au/, 2020.
Accessed: 2020-10-24.

[279] T. PIKETTY AND A. GOLDHAMMER, *Capital in the Twenty-First Century*, Harvard University Press, Mar. 2014.

[280] M. POLETAEV AND C. ROBINSON, *Human capital specificity: evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000*, Journal of Labor Economics, 26 (2008), pp. 387–420.

[281] M. PRIES AND R. ROGERSON, *Hiring policies, labor market institutions, and labor market flows*, Journal of Political Economy, 113 (2005), pp. 811–839.

[282] PRODUCTIVITY COMMISSION, *Aspects of Structural Change in Australia*, tech. rep., Australian Government, 1998.

[283] ——, *Rising inequality? A stocktake of the evidence*, tech. rep., Australian Government, Aug. 2018.

[284] G. QUINTINI, *Right for the Job*, tech. rep., OECD, 2011.

[285] D. RADCLIFFE, *Covid-19 has ravaged american newsrooms–here‚Äôs why that matters*, Available at SSRN 3693903, (2020).

[286] M. REYNOLDS, *DeepMind's neural network teaches AI to reason about the world*, New Scientist, (2017).

[287] S. RICHARDSON, *What is a skill shortage?*, tech. rep., National Centre for Vocational Education Research, 2007.

[288] M. RICKETSON, A. DODD, L. ZION, AND M. WINARNITA, *"Like being shot in the face" or "I'm glad I'm out": Journalists' experiences of job loss in the Australian media industry 2012–2014*, Journalism Studies, 21 (2020), pp. 54–71.

[289] C. ROBINSON, *Occupational mobility, occupation distance, and specific human capital*, Journal of Human Resources, 53 (2018), pp. 513–551.

[290] U. M. RODRIGUES AND Y. PARADIES, *News Consumption Habits of Culturally Diverse Australians in the Digital Era: Implications for Intercultural Relations*, Journal of Intercultural Communication Research, 47 (2018), pp. 38–51.

[291] E. M. ROGERS, *New product adoption and diffusion*, Journal of Consumer Research, 2 (1976), pp. 290–301.

[292] S. RUSSELL AND P. NORVIG, *Artificial Intelligence: A Modern Approach*, Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd ed., 2009.

[293] S. L. RYNES, R. D. BRETZ, JR, AND B. GERHART, *The importance of recruitment in job choice: A different way of looking*, Personnel Psychology, 44 (1991), pp. 487–521.

[294] Z. SAMIOS, *Investors look to salvage parts of AAP as newswire faces closure*. https://www.smh.com.au/business/companies/investors-look-to-salvage-parts-of-aap-as-newswire-faces-closure-20200302-p54627.html, Mar. 2020. Accessed: 2020-05-01.

[295] M. U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, Harvard Journal of Law and Technology, 29 (2016), pp. 353–400.

[296] T. W. SCHULTZ, *Investment in Human Capital*, The American Economic Review, 51 (1961), pp. 1–17.

[297] J. A. SCHUMPETER, *Capitalism, Socialism, and Democracy: Third Edition*, Harper Perennial Modern Classics, unknown edition ed., Nov. 2008.

[298] K. SCHWAB, *Accelerating Workforce Reskilling for the Fourth Industrial Revolution: An Agenda for Leaders to Shape the Future of Education, Gender and Work*, tech. rep., World Economic Forum, July 2017.

[299] ——, *The Fourth Industrial Revolution*, World Economic Forum, Jan. 2017.

[300] K. SCHWAB AND R. SAMANS, *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*, tech. rep., World Economic Forum, Jan. 2016.

[301] J. SCOTT ARMSTRONG, *Long-Range Forecasting: From Crystal Ball to Computer*, Wiley-Interscience, 2 edition ed., July 1985.

[302] K. L. SHAW, *A formulation of the earnings function using the concept of occupational investment*, Journal of Human Resources, (1984), pp. 319–340.

[303] ——, *Occupational change, employer change, and the transferability of skills*, Southern Economic Journal, (1987), pp. 702–719.

[304] M. SHERWOOD AND P. O'DONNELL, *Once a Journalist, Always a Journalist?*, Journalism Studies, 19 (2018), pp. 1021–1038.

[305] C. SHIRKY, *Newspapers and thinking the unthinkable*, Risk Management, 56 (2009), p. 24.

[306] J. N. SHURKIN, *Engines of the Mind: The Evolution of the Computer from Mainframes to Microprocessors*, Norton, 1996.

[307] S. T. SHUTTERS, R. MUNEEPEERAKUL, AND J. LOBO, *Constrained pathways to a creative urban economy*, Urban Stud., 53 (2016), pp. 3439–3454.

[308] U. SILA, *Job displacement in Australia: Evidence from the HILDA survey*, tech. rep., OECD, 2019.

[309] D. SILVER, T. HUBERT, J. SCHRITTWIESER, I. ANTONOGLOU, M. LAI, A. GUEZ, M. LANCTOT, L. SIFRE, D. KUMARAN, T. GRAEPEL, T. LILLICRAP, K. SIMONYAN, AND D. HASSABIS, *A general reinforcement learning algorithm that masters chess, shogi, and go through self-play*, Science, 362 (2018), pp. 1140–1144.

[310] V. SMIL, *Transforming the Twentieth Century: Technical Innovations and Their Consequences*, Oxford University Press USA, illustrated edition ed., May 2006.

[311] I. STEWART, D. DE, AND A. COLE, *Technology and People: The great job-creating machine*, tech. rep., Deloitte, 2015.

[312] J. E. STIGLITZ, *The Great Divide: Unequal Societies and What We Can Do About Them*, W. W. Norton & Company, 1 edition ed., Apr. 2016.

[313] ——, *Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution*. Sept. 2017.

[314] J. E. STIGLITZ AND B. C. GREENWALD, *Externalities in Economies with Imperfect Information and Incomplete Markets*, Quarterly Journal of Economics, 101 (1986), pp. 229–264.

[315] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 2016.

[316] E. Tang, *Australia has experienced the longest period of economic growth in the developed world*.
https://www.austrade.gov.au/News/Economic-analysis/australia-has-experienced-the-longest-economic-growth-among-major-developed-world, Mar. 2017.
Accessed: 2018-9-11.

[317] J. Tang and W. Wang, *Product market competition, skill shortages and productivity: Evidence from Canadian manufacturing firms*, Journal of Productivity Analysis, 23 (2005), pp. 317–339.

[318] C. Taylor, J. Carrigan, H. Noura, S. Ungur, J. van Halder, and G. S. Dandona, *Australia's automation opportunity: Reigniting productivity and inclusive income growth*, tech. rep., McKinsey & Company, 2019.

[319] S. J. Taylor and B. Letham, *Forecasting at scale*, The American Statistician, 72 (2018), pp. 37–45.

[320] J. Y. Thong, *An integrated model of information systems adoption in small businesses*, Journal of management information systems, 15 (1999), pp. 187–214.

[321] R. H. Topel and M. P. Ward, *Job mobility and the careers of young men*, The Quarterly Journal of Economics, 107 (1992), pp. 439–479.

[322] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou, *Fixing the train-test resolution discrepancy: FixEfficientNet*, (2020).

[323] M. Trajtenberg, *Artificial Intelligence as the Next GPT: A Political-Economy Perspective*, in The Economics of Artificial Intelligence: An Agenda, University of Chicago Press, Jan. 2018, pp. 175–186.

[324] A. M. Turing, *Computing Machinery and Intelligence*, Mind, 59 (1950), pp. 433–460.

[325] University of Technology Sydney, *Graduate Research Candidature Management, Thesis Preparation and Submission Procedures*.

https://www.uts.edu.au/sites/default/files/2020-10/GRS-Graduate-Research-Candidature-Management-Thesis-Preparation-and-Submission-Procedures-V1.10.pdf, 2020.
Accessed: 2020-12-13.

[326] U.S. DEPARTMENT OF LABOR, *O*NET*.
https://www.onetonline.org/.
Accessed: 2020-9-22.

[327] US EXECUTIVE OFFICE OF THE PRESIDENT, *The National Artificial Intelligence Research and Development Strategic Plan*, tech. rep., National Science and Technology Council. Networking and Information Technology Research and Development Subcommittee., Oct. 2016.

[328] M. VAN DER MERWE, *Factors Affecting an Individual's Future Labour Market Status*, RBA Bulletin, (2016), pp. 11–22.

[329] S. VILLE AND G. WITHERS, *The Cambridge Economic History of Australia*, Cambridge University Press, Oct. 2014.

[330] T. L. VOLLRATH, *A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage*, Weltwirtsch. Arch., 127 (1991), pp. 265–280.

[331] F. VONA AND D. CONSOLI, *Innovation and skill dynamics: a life-cycle approach*, Industrial and Corporate Change, 24 (2015), pp. 1393–1415.

[332] C. WAHLQUIST, *Australian Associated Press sold to consortium of investors and philanthropists at 11th hour*, 2020.
Accessed: 2020-10-22.

[333] S. WALDMAN AND C. SENNOTT, *The coronavirus is killing local news*, The Atlantic, (2020).

[334] M. M. WALDROP, *More from Moore*, Nature, 520 (2015), p. 408.

[335] E. WASMER, *General versus specific skills in labor markets with search frictions and firing costs*, American Economic Review, 96 (2006), pp. 811–831.

[336] N. WATSON AND M. WOODEN, *Sample attrition in the HILDA survey*, Australian Journal of Labour Economics, 7 (2004), pp. 293–308.

[337] D. WILDING, P. FRAY, S. MOLITORISZ, AND E. MCKEWON, *The impact of digital platforms on news and journalistic content*, tech. rep., University of Technology Sydney, 2018.

[338] M. WILLENS, *The second wave of media layoffs is here*, 2020.
Accessed: 2020-10-22.

[339] WORKPLACE GENDER EQUALITY AGENCY, AUSTRALIAN GOVERNMENT, *Equal Pay Day*.
https://www.wgea.gov.au/terms/equal-pay-day.
Accessed: 2020-11-25.

[340] WORLD ECONOMIC FORUM, *The Global Risks Report 2017*, Tech. Rep. 12, World Economic Forum, 2017.

[341] ——, *Towards a Reskilling Revolution A Future of Jobs for All*, tech. rep., World Economic Forum and The Boston Consulting Group, 2018.

[342] ——, *Towards a Reskilling Revolution: Industry-Led Action for the Future of Work*, tech. rep., World Economic Forum and The Boston Consulting Group, 2019.

[343] WORLD HEALTH ORGANIZATION, *Ageing and health*.
https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, Feb. 2018.
Accessed: 2020-2-8.

[344] WORLD INEQUALITY DATABASE, *Australia – WID*.
https://wid.world/country/australia/, 2020.
Accessed: 2020-8-9.

[345] S. YOUNG AND A. CARSON, *What is a journalist? The view from employers as revealed by their job vacancy advertisements*, Journalism Studies, 19 (2018), pp. 452–472.

[346] Y. ZHOU, T. LI, J. SHI, AND Z. QIAN, *A CEEMDAN and XGBOOST-Based approach to forecast crude oil prices*, Complexity, 2019 (2019).

[347] L. ZION, M. SHERWOOD, P. O'DONNELL, A. DODD, M. RICKETSON, AND T. MARJORIBANKS, *'It has a bleak future': The effects of job loss on regional and rural journalism in Australia*, (2016).

[348] L. ZION, M. SHERWOOD, P. O'DONNELL, T. MARJORIBANKS, M. RICKETSON, A. DODD, AND M. WINARNITA, *New Beats report: mass redundancies and career change in Australian journalism*, tech. rep., Jan. 2018.