NOVA IMS
Information Management School

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

AI-Powered Career Navigation Platform

Predicting Automation Risk and Guiding Professionals to Optimal Occupation Transitions with a Focus on Generation X in Portugal

Ricardo da Costa Almeida

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

AI-powered Career Navigation Platform

Predicting Automation Risk and Guiding Professionals to Optimal Occupation Transitions with a Focus on Generation X in Portugal

by

Ricardo da Costa Almeida

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialisation in Data Science.

Supervised by

Roberto Henriques, PhD in Information Management, NOVA IMS

June, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, June 24, 2024

DEDICATION

I dedicate this dissertation to my beloved family, Raquel Pereira, for their unwavering support, and to my friends for making this journey easier by assisting me in every way they could. Special thanks to my esteemed colleagues, Ianis Rușitoru and Nichita Zamisnii, whose support and companionship have been invaluable throughout our two-year journey.

ACKNOWLEDGEMENTS

I express my deepest gratitude to my supervisor, Roberto Henriques, for his invaluable guidance and insightful support throughout this research project. His expertise and dedication have been instrumental in helping me achieve my best performance. Thank you for your unwavering support and encouragement.

ABSTRACT

The rapid advance of artificial intelligence (AI) and automation technologies has raised significant concerns about job displacement, particularly among Generation X professionals in Portugal. This project addresses these concerns by developing an interactive recommendation system named Career Xplorer, designed to help professionals navigate potential career transitions in a personalised way. Using advanced Natural Language Processing (NLP) models such as BERT and GPT-4 and a task-based approach that assesses automation risk for individual tasks rather than entire occupations, Career Xplorer estimates the automation probability for each user's occupation. The recommendation system also suggests alternative occupations with lower automation risks and provides a detailed job transition guide outlining the essential skills and competences needed for these transitions. This guide helps users identify which skills to focus on and which to deprioritise, facilitating a more confident and smoother professional transition. The study finds that not only occupations with repetitive and low-skilled tasks face the greatest risks of automation but also more technical roles, mainly due to the presence of well-established rules. On the other hand, more communicative and creative occupations have lower levels, suggesting an area of greater security in a labour market impacted by AI and automation. By empowering Generation X professionals with the tools and knowledge to proactively manage their careers amidst the evolving technological landscape, Career Xplorer helps mitigate the impacts of job displacement and promotes long-term career stability and growth while taking employee happiness into account.

KEYWORDS

Job Displacement; Job Transition Guide; Automation Risk; Natural Language Processing; Interactive Recommendation System

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
BERT	Bidirectional Encoder Representations for Transformers
DOT	Dictionary of Occupational Titles
EQF	European Qualifications Framework
ESCO	European Skills, Competences, Qualifications and Occupations
EU	European Union
GPT	Generative Pre-training Transformer
INE	National Statistics Institute
LLM	Large Language Model
MDS	Multidimensional Scaling
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
OER	Open Education Resources
O*NET	Occupational Information Network
PIAAC	Programme for the International Assessment of Adult Competences
SD	Standard Deviation
USA	United States of America

1. INTRODUCTION

1.1. BACKGROUND

In the face of technological advances, in recent years, many aspects of daily life have been the scene of significant changes, giving wings to a transformative technological wave spread worldwide (Nair et al., 2021). The discipline that had the most impact since its inception stands out: artificial intelligence (AI). From its affirmation as an academic field at the Dartmouth Conference in 1956 to the first steps taken in fields such as machine learning and neural networks during the 1980s and 1990s, and more recently the emergence of innovative natural language processing models such as GPT-3 and GPT-4, as well as and computer vision, the path of AI reflects a consistent and exponential evolution, having been affirmed over time (Haenlein & Kaplan, 2019; Benbya et al., 2020).

Given that this new technological era shows no signs of slowing down, keeping abreast of its evolution requires a high level of understanding, interaction, and time, which is often impossible to reconcile with the rhythm of our lives. In some fields, AI already outperforms humans, such as in image recognition and reading comprehension, so it's crucial to adapt but also to analyse the impact of the implications these changes will have on the future of each of us so that there are fewer setbacks and more benefits (Giattino et al., 2023).

In addition to its undeniable advantages, such as improving resource allocation, increasing productivity and efficiency at work in numerous sectors, and contributing to social welfare, bringing improvements in health, education, and transportation, AI also brings together a series of challenges such as increased technological dependence, economic inequality, the distinction between ethically correct and incorrect and the partial or total replacement of human beings by machines or computers, in the performance of tasks associated with different professional careers, the latter being the central focus of our analysis (Graetz & Michaels, 2018; Tai, 2020; World Economic Forum, 2024). That said, as the loss of jobs is a real risk in some sectors, namely in those that have a more significant number of repetitive and low-skilled tasks, understanding the impact of this new technological wave is essential, not least because the consequences of it will have a different effect on the lives of each of us, namely concerning different generational groups, which have distinct characteristics (Brynjolfsson e McAfee, 2011, 2014; World Economic Forum, 2023). Among these, Generation X professionals stand out, as they are at the centre of concerns about transformation or job loss.

Generation X, made up of individuals born between 1965 and 1980 (Dimock, 2019), is at the peak of their careers. However, they face a demanding personal life. Being characterised by numerous family obligations, including responsibility for their children and sometimes even their parents, as well as financial concerns such as managing their salary in the face of unpaid debts and given the high divorce rates recorded in recent years, it is understandable why this generation is in an alarming zone regarding the risk of job loss, which comes from the possibility of their jobs being replaced by AI or the difficulty of adapting to the integration of AI into some work tasks (Steiner e Fletcher, 2017; PORDATA, 2022). In the workplace, Gen Xers value a healthy work-life balance. They have a very autonomous work methodology, making them more individualistic than previous generations and, in this way, they may face an increased difficulty in the learning phase of more advanced and challenging technological fields, such as AI, making them less susceptible to transitions to new and unknown territories (Jorgensen 2003; Fishman, 2016). In addition, as Portugal is the focus of this study, this generation does not have an age range precisely close to the Portuguese retirement age, which is currently approximately 67 years, which means that these individuals are between 7 and 22 years old until they reach retirement age, which further aggravates their situation.

Besides the demographic features that characterize the generation under study, it is essential to note that Portugal lags behind the rest of Europe and is not an exception when it comes to the development and adoption of AI (Fonseca et al., 2018; Brattberg et al., 2020; Małkowska et al., 2021). Additionally, Portugal is often absent from relevant studies that have used primary data sources, such as the Programme for the International Assessment of Adult Competences (PIAAC), where Portugal is not included (Nedelkoska and Quintini, 2018). The consequences of AI integration in the labour market are already visible throughout Europe, especially in the most advanced countries in this matter. This suggests that when Portugal overcomes its technological gap compared to other European countries, the negative impacts, such as job loss or transformation, will be greater if there is no prior and solid preparation. Therefore, the geographical restriction defined for this project, driven by both patriotism as it is the native country of those responsible for the project, and the need to study a specific labour market, acknowledges the unique conditions and challenges that shape the experiences of the subjects under analysis. In this way, the research aims to provide insights and guidance based on data that are not only timely but also deeply pertinent to professionals who belong to Generation X and work in the Portuguese labour market.

In short, the project aspires to provide these professionals with the tools and knowledge to proactively navigate their career paths in an unpredictable era profoundly shaped by AI and automation. This effort aims to ensure that they are well prepared for the future, mitigating undesirable surprises and alleviating the possible complexities that may be seen in career transitions, during their already overloaded lives.

1.2. Problem Definition and Project Work Aim

Given the impacts driven by the adoption and integration of AI technologies throughout the European labour market, this project seeks to respond to the need for proactive career guidance and support mechanisms for Generation X professionals in Portugal (Lane et al., 2023). From these challenges, our central research questions arise:

- What are the potential impacts of AI and automation on occupations for Generation X professionals in Portugal, and what are the expected and optimal occupational transitions?
- In what ways can a recommendation system assist Generation X professionals in Portugal in proactively managing and planning their careers in response to AI and automation trends?

In response to the questions and to provide reliable and detailed preparation for Generation X professionals in Portugal, we propose to develop a recommendation system and integrate it into a web application, resulting in an innovative platform. This platform will allow a personalised analysis, based on actual data, of the risk of automation across various occupations in the labour market. Moreover, it will present a guide of skills and competences to facilitate transitions to better jobs chosen by users, considering the level of exposure to automation risk and similarity.

Given that automation varies by industry and occupational structure (Nedelkoska and Quintini, 2018), our approach aims to reduce the impact of task automation in Portugal, regardless of AI adoption and integration levels in the Portuguese labour market. That said, we will present a standard structure of each occupation, including constituent tasks and levels of frequency and importance, at the European level (ESCO), allowing users to edit the tasks they perform, as well as their importance and frequency. This customisation will cover the automation risk level for each individual's specific role.

With this centralised information, Generation X professionals can identify which tasks will be impacted by AI within their daily work and adapt accordingly. Some tasks may change in execution, supported by AI rather than being entirely replaced, thus necessitating employee training (Benbya et al., 2020). If adaptation is not in their best interest, the platform also prepares users for a smooth transition to a similar occupation of their choice by identifying the relevant skills and competences needed for the transition.

1.3. OBJECTIVES

The objective of this project is twofold: to upgrade the ESCO (European Skills, Competences, Qualifications and Occupations) database and to build an interactive recommendation system that can support Generation X professionals in dealing with the impact of AI on their work environment, particularly due to the automation of processes previously carried out exclusively by humans. This dual aim promotes a more informed present and a more promising future. Focusing on Portugal's three largest Generation X major work groups, this research seeks to elucidate the following key aspects.

Firstly, to enhance the ESCO database, we aim to predict and include the importance and frequency of tasks within occupations - features currently available in the O*NET (Occupational Information Network) database but missing in ESCO. By fine-tuning a pre-trained BERT model to predict importance and frequency of tasks based on the names and descriptions of occupations in ESCO and the descriptions of tasks in O*NET, we ensure the ESCO database becomes more comprehensive and useful for the analysis and development of future European studies.

Secondly, for users to understand the impact of artificial intelligence on a given occupation, namely how likely it is to be automated, we calculate the risk of automation that the tasks that constitute it present. In this way, in addition to measuring and providing an analysis of the risk of automation in the workplace, it is possible to identify which tasks are likely to be entirely replaced by AI and which are most likely to be transformed, resulting in a collaboration between the worker and AI. In this context, it is easier to allocate training to an employee who is unfamiliar with the new work methodology.

Thirdly, we developed a recommendation system based on the transition between similar occupations, where the final recommendation is characterised by a set of skills and competences organised by their relevance and distance from each other. This information, when centralised, allows Generation X professionals to prepare for a safer transition, knowing which skills they should learn, which they should retain, and which they can leave in the background because they have already consolidated them.

Next, we provided a detailed analysis of expected and optimal transitions, highlighting both natural career progression pathways and strategic career moves with high skill similarity. This analysis serves to guide Generation X professionals in identifying potential career shifts that align closely with their existing skills, thereby minimising the need for extensive retraining. By presenting these transitions, we empower workers to make informed decisions about their career trajectories, ensuring they can adapt more effectively to the evolving demands of the labour market and mitigate the risks associated with automation.

Finally, to provide utility and extract value from the entire developed system, we created a web application that serves as an interface, unlocking the interactivity between the user and the developed system. In addition, this application also offers a personalised experience to the user, allowing them to receive an occupation automation risk assessment based on their specific situation and choose the occupation to which they want to move, depending on their personal and professional reasons, directly influencing the recommendations they will receive.

1.4. SUCCESS CRITERIA

In assessing the success and accomplishment of the outlined research objectives, specific metrics and indicators will be employed to determine the effectiveness of the research objectives.

The first criterion for success will be measured by conducting a comprehensive analysis of job industry trends within the Portuguese context. This involves identifying the three largest major groups, or 1-digit ISCO codes, for Generation X in Portugal.

The second criterion focuses on the successful upgrade of the ESCO database by incorporating the importance and frequency of tasks within occupations. This will be measured by the performance metrics and consistency of the predictions made by the fine-tuned pre-trained BERT model and the extent to which these additions enhance the comprehensiveness and usability of the ESCO database for future European studies.

The third criterion focuses on the consistency of the risks of automation in different occupations. This prediction will be based on empirical data and alignment with the results of similar studies, ensuring a reliable assessment of the professional functions likely to be displaced by AI and automation.

The fourth criterion evaluates the quality of the recommendations provided by the recommendation system. Success in this aspect will be determined by the coherence, validity, and practical interest of the recommended career paths. This involves ensuring that the recommendations are logically consistent, grounded in real-world applicability, and theoretically appealing to the target users.

Lastly, the fifth criterion involves the successful implementation of an interactive platform designed to provide accessible and personalised career information to users. Its success will be evaluated based on its functionality, user interface, and the extent to which it meets the specific informational needs of Generation X professionals in Portugal.

These success criteria collectively form the evaluative framework for the research objectives, ensuring that the research outcomes are informative, insightful, and practically beneficial for the target demographic.

1.5. Study Impact on Relevance

This project aims to increase job security by supporting professionals in learning skills and competences aligned with their interests and needs, expanding career opportunities in the most common occupations for Generation X professionals in Portugal, and promoting personal and professional growth. In addition, detailed skills recommendation encourages individuals to pursue rewarding careers by aligning their development with their passions.

A significant advance in this study was reformulating the automation risk calculations for occupations, incorporating the frequency of tasks and their importance. This new calculation methodology makes the analysis closer to reality, providing a more detailed understanding of the impact of automation on different occupational roles.

Another distinctive aspect of this research is integrating a recommendation system into a web application characterised by a high level of personalisation, providing an analysis tailored to each user's situation. The design of this platform allows for future scalability to any European country and the inclusion of occupations, benefiting every generation. It serves as a valuable tool for career planning and transitions across Europe, addressing the challenges posed by AI and automation in the modern labour market. By incorporating data from specific labour markets and relevant indicators, the system offers substantial value to local societies. This versatility ensures that the platform can be adapted to different demographics and regional needs.

As a consequence of this process, we also upgrade the ESCO database to include the importance and frequency of tasks within occupations, making this data source more robust and encouraging future studies to utilise ESCO instead of adapting O*NET. This enhancement ensures that demographic-focused data is harnessed effectively, supporting European-centric research and providing a more accurate reflection of the European labour market.

Finally, for more technologically backward countries, such as Portugal, this system offers additional advantages. Beyond personalised career support, it mitigates the potential impacts of AI on professional lives, preparing individuals for future changes through informed policies and labour market initiatives. By providing a solid foundation for adapting to the evolving job landscape, this project not only benefits individuals but also contributes to the overall resilience and adaptability of the workforce.

1.6. Outline

Regarding the project's structure, Section 1 is the introduction, followed by five main sections. Section 2 is the literature review, beginning with an overview of O*NET and ESCO, defining key terms and concepts essential for the subsequent analysis and findings (Section 2.1). It covers various methodological approaches to automation risk assessment (Section 2.2) and reviews studies on recommendation systems and their applications in career guidance and career transition support (Section 2.3). Section 2.4 provides a summary of the entire literature review, encapsulating the key insights and frameworks discussed.

Section 3, the methodology, follows the CRISP-DM framework, detailing the research, data collection methods, and analytical approaches used to develop the interactive recommendation system. It includes Business Understanding (Section 3.1), Data Understanding & Preparation (Section 3.2), Modelling (Section 3.3), which is divided into Automation Risk at the Occupation Level (Section 3.3.1) and Similarity between Occupations and Skills and Competences (Section 3.3.2), Evaluation (Section 3.4), and Deployment (Section 3.5).

Section 4 presents the results and discussion, including the results of the importance and frequency indicators obtained by fine-tuning the pre-trained BERT model (Section 4.1), the automation risk calculations at occupation level (Section 4.2), the analysis of similarities across skills and occupations through cluster analysis, examining the relationships between skills and competences to facilitate potential career transitions (Section 4.3), the development of a recommendation system that suggests similar occupations with lower automation risk and provides a skill guide to support a smooth transition based on categorised skill distance and importance (Section 4.4), the analysis of expected and optimal transitions between occupations based on skill similarity, highlighting career pathways from high-risk to lower-risk occupations and ensuring smoother transitions with minimal retraining (Section 4.5), and the creation of a web application to provide an interactive platform for users to explore career transitions, customise their occupation structure, and receive personalised recommendations for skill development and career planning (Section 4.6).

Section 5 is the conclusion, summarising the main findings and contributions of the project to career guidance and AI impact assessment. Section 6 discusses limitations and future implementations, providing recommendations for future research and practical applications. The appendix includes supplementary material, such as graphs illustrating the recommendation system process and results and images of the developed platform interface.

2. LITERATURE REVIEW

In this chapter, we look at the most impactful methodologies for assessing the risk of automation in the labour market. We begin by defining key terms related to O*NET and ESCO (Section 2.1). Next, we highlight the importance of an approach based on tasks rather than occupations, the use of advanced language models with various Natural Language Processing (NLP) capabilities for measure the automation risk of different occupations, and the choice of a data source aimed directly at the reality of Portugal (Section 2.2). Finally, we explore recommendation systems focused on labour transitions, emphasising the importance and difference of having an interactive dimension (Section 2.3).

2.1. O*NET AND ESCO: TERMINOLOGY

In the field of labour market analysis, professional information systems such as O*NET and ESCO play an essential role. These systems provide data on various aspects of careers that are crucial for labour force growth, policy creation and academic research. In this section, we will explore the background, structures and uses of O*NET and ESCO, while defining terms such as occupations, tasks, skills and competences. This will help to understand the analysis carried out throughout this project, which consists of the development of Career Xplorer, an AI-driven platform designed to help Generation X professionals in Portugal with career transitions while facing automation challenges.

2.1.1. O*NET: Overview and Structure

2.1.1.1. History and Development

Introduced in the late 1990s, the Occupational Information Network (O*NET) was developed by the US Department of Labor to replace a previous system that listed job descriptions and requirements, known as the Dictionary of Occupational Titles (DOT). In this way, O*NET provides a more extensive and detailed database of occupational information. This database is updated several times a year through surveys and expert contributions, thereby supporting job seekers, employers, and researchers with reliable and up-to-date data (Peterson et al., 2001; Hilton & Tippins, 2010).

2.1.1.2. Structure and Components

O*NET's structure is based on the O*NET Content Model, which includes six major domains: Worker Characteristics, Worker Requirements, Experience Requirements, Occupational Requirements, Workforce Characteristics, and Occupation-Specific Information. Each domain contains detailed descriptions that capture the essential aspects of occupations, such as skills, abilities, knowledge, work activities, and interests (National Center for O*NET Development, 2024; Burrus et al., 2013). Worker Characteristics refer to attributes that influence a person's capacity to perform work tasks, including abilities, interests, and work values. Worker Requirements focus on the skills and knowledge required for specific occupations. Experience

Requirements cover the experience and training necessary for job performance. Occupational Requirements detail the work activities and context of occupations. Workforce Characteristics provide labour market information and trends, while Occupation-Specific Information includes unique details pertinent to occupations.

2.1.1.3. Data Collection and Usage

O*NET data is collected through surveys and expert assessments, ensuring that the information is comprehensive and current. This data is widely used in career guidance, workforce development, policy-making and academic research (Jenkins & Moses, 2014; Crouter et al., 2006), supporting multiple stakeholders in making informed decisions about employment and training (National Centre for O*NET Development, 2024).

2.1.2. ESCO: Overview and Structure

2.1.2.1. History and Development

The European Skills, Competences, Qualifications and Occupations (ESCO) classification was developed by the European Commission. Its purpose is to bridge the gap between education and the labour market by providing a common language for skills, competences, qualifications, and occupations across Europe (ESCO, 2024). Launched in 2013, ESCO supports several EU initiatives, including the EQF (European Qualifications Framework), which provides a reference framework for comparing qualifications across different European countries, and Europass, a set of documents that help individuals present their skills and qualifications in a standardised format (ESCO, 2024; Smedt et al., 2015).

2.1.2.2. Structure and Components

ESCO's classification system is divided into three main pillars: Occupations, Skills/Competences, and Qualifications. Each pillar is designed to provide detailed information to support labour market and educational needs (ESCO, 2024; le Vrang et al., 2014). The Occupations pillar includes a list of job titles and descriptions, organised in a hierarchical structure. ESCO classifies occupations into 9 major groups, also known as 1-digit ISCO codes. The International Standard Classification of Occupations (ISCO) is a system developed by the International Labour Organization (ILO) to classify occupational information internationally (International Labour Organization, 2012). The 9 major groups in ESCO are Managers, Professionals, Technicians and Associate Professionals, Clerical Support Workers, Service and Sales Workers, Skilled Agricultural, Forestry and Fishery Workers, Craft and Related Trades Workers, Plant and Machine Operators and Assemblers, and Elementary Occupations. The Skills/Competences pillar provides descriptions of the skills and competences required for different occupations, including both hard and soft skills. The Qualifications pillar offers information about qualifications and certifications relevant to specific occupations and skills.

2.1.2.3. Data Collection and Usage

ESCO data is compiled from multiple sources, including national classification systems, educational institutions, and organisations from different sectors, and is continuously updated to reflect possible changes in the labour market and educational standards. ESCO is used by employers, job seekers, policymakers, and educational institutions to facilitate their migration and alignment between skills supply and demand (ESCO, 2024; Smedt et al., 2015).

2.1.3. Definitions of Key Terms

2.1.3.1. Occupations

Occupations are defined differently by O*NET and ESCO. In O*NET, occupations are defined as a set of activities or tasks that individuals are paid to perform as part of their employment. Each occupation in O*NET is described by its tasks, knowledge, skills, abilities, and work context (National Center for ONET Development, 2024). In contrast, ESCO defines an occupation as a group of jobs requiring similar tasks and specialised knowledge, organised hierarchically to show relationships between roles and levels of responsibility (ESCO, 2024). While both O*NET and ESCO provide detailed descriptions of occupations, O*NET focuses more on the specific attributes and requirements of each occupation, whereas ESCO provides a broader classification system that links occupations to skills and qualifications.

2.1.3.2. Tasks

Tasks, according to O*NET, are the specific activities that employees perform as part of their occupations. These tasks are detailed and include information on the context, frequency of performance, importance and relevance within the occupation they belong to (National Center for O*NET Development, 2024). In ESCO, tasks are described as the core activities associated with each occupation, providing a clear understanding of role expectations and encompassing the individual activities that make up the work performed in each occupation (ESCO, 2024). Both O*NET and ESCO provide detailed task descriptions, though O*NET offers more granularity by including context, frequency, importance and relevance information.

2.1.3.3. Skills

Skills are defined by O*NET as basic skills (fundamental abilities) and cross-functional skills (transferable skills applicable to various jobs), which include cognitive skills like problem-solving and psychomotor skills like coordination (National Center for ONET Development, 2024). In contrast, ESCO defines skills as encompassing both hard skills (technical abilities) and soft skills (interpersonal abilities), emphasising the competencies required to perform job tasks effectively, and the abilities to perform tasks and solve problems (ESCO, 2024). While O*NET categorises skills into basic and cross-functional, ESCO uses the hard and soft skills classification. Both systems aim to provide a comprehensive overview of the skills required for different occupations.

2.1.3.4. Competences

Competences in O*NET are often described in terms of knowledge, skills, and abilities that enable performance in a job. These are aligned with the requirements and tasks of the occupation (National Center for ONET Development, 2024). ESCO defines competences as the combination of skills, knowledge, and abilities required to perform occupational tasks effectively, closely linked to qualifications and educational outcomes, encompassing the skills, knowledge, and attitudes necessary to perform tasks to a certain standard (ESCO, 2024). Both O*NET and ESCO view competences as a combination of attributes necessary for job performance, with ESCO placing a stronger emphasis on linking competences with qualifications.

2.1.4. The Crosswalk Between O*NET and ESCO

The integration of O*NET and ESCO data is facilitated by a crosswalk that links occupations between the two systems. This cross-referencing is available in both domains and is essential for leveraging the strengths of both databases, ensuring consistency and relevance for studies in this field.

According to the O*NET-ESCO Technical Report, this crosswalk leverages machine learning and natural language processing to map the classifications, supplemented by human validation to ensure accuracy and coherence (European Commission, 2022).

2.2. Methodological Approaches to Automation Risk Assessment

2.2.1. Occupation-Based vs. Task-Based Approach

The possible automation of labour market tasks began to raise significant concerns about a decade ago, leading to detailed analyses and expert studies (Frey and Osborne, 2013; Arntz et al., 2016). Pioneering work by Frey and Osborne (2013) drew attention to the potential impact of automation on the labour market through an assessment of professional sectors. Their study offered a comprehensive view of the challenges posed by automation while ignoring the heterogeneity of the tasks that make up jobs, assuming that two people with the same job performed the same tasks. This approach revealed that almost half, around 47%, of jobs in the United States would be at high risk of automation, which sparked discussions and debates about the possible displacement of jobs and the reshaping of the labour landscape.

This huge and frightening percentage has also led other experts to interpret these results as somewhat exaggerated figures, proposing that the approach adopted was inappropriate. This was the case with Arntz et al. (2016), who, in response to the potential overestimation of job displacement, presented an approach centred on the difference in tasks within the same jobs, countering the homogeneity assumed by Frey and Osborne. Their analysis reinforced the importance of a more detailed assessment for more realistic forecasts, obtaining a substantially lower risk of automation, indicating that only around 9% of jobs in the

Organisation for Economic Co-operation and Development (OECD) countries were highly susceptible to automation.

Since then, the contrast between the Occupation-Based Approach and the Task-Based Approach has become clear since “workers' task structures differ remarkably within occupations (Autor and Handel 2013). Thus, even within occupations, workers are likely to be exposed very differently to automation, depending on the tasks they perform” (Arntz et al., 2016, p. 12).

Building on the work of Arntz et al. (2016), Nedelkoska and Quintini (2018) further refined the Task-Based Approach methodology using data from PIAAC, namely the Adult Skills Survey. This survey aims to provide a comprehensive overview of the key competences of working populations, how they are used and changed by work, education and learning, and their impacts on wages, employment, the economy and social welfare (Martin, 2018). In this way, the authors provided a more detailed assessment of the risk of automation in different occupations in 32 different member countries of the OECD. Their findings suggest that although the overall risk of automation is lower than estimates previously made at the occupation level, it is still significant, with around 14% of jobs in the countries under study at high risk of being automated.

In addition, they concluded that the risk of automation is not evenly distributed across occupations, with lower-paid and low-skilled jobs being disproportionately affected. This highlights the need for and focus of future work on the development of retraining and skills improvement programmes to help workers in these professions adapt to the new demands of the labour market, thus justifying the development of a skills and competences guide focused on future professional transitions, and consequently one of the motivations for our project.

Also based on the work of Arntz et al. (2016) but using a more specific analysis compared to Nedelkoska and Quintini (2018), Lassébie and Quintini (2022) focused on identifying which specific skills and competences automation technologies can replicate. Using metrics from the O*NET entity and expert knowledge, they concluded, with the help of the EU Labour Force Survey, that on average, in OECD countries, the occupations most at risk of automation account for around 28% of employment. Compared to previous studies, this percentage is higher, justified by technological advances and differences in the methodologies applied. Lassébie and Quintini (2022) also point out that tasks previously seen as bottlenecks (i.e. tasks that cannot be automated with the existing technologies) in previous studies are beginning to be automatable, such as knowledge in the field of fine arts and various psychomotor tasks, marking a significant change in the impact of automation on the labour market. According to their findings, the sectors most at risk of automation include construction and extraction, farming, fishing, and forestry, production, and transportation and material moving occupations. In contrast, the sectors least at risk of automation cover community and social service, management, educational instruction and library, and legal occupations.

2.2.2. Automation Risk Assessment

Like Dawson et al. (2021), who used several Australian job advertisements to create an indicator of AI in the Australian labour market, Acemoglu et al. (2022) analysed the impact of AI on the USA (United States of America) labour market, using detailed data from online job postings since 2010. It found a significant growth in AI-related vacancies, especially after 2015, in establishments with high exposure to AI, where new technologies were part of the job structure. These establishments reduced hiring for non-AI-related positions, suggesting that AI is replacing humans in specific tasks. Although the authors did not find a clear relationship between AI exposure and employment or wage growth at the occupational or industrial level, they did identify a marked increase in AI-related activity, a shift in the skills required by AI-exposed establishments and a reduction in overall hiring and non-AI-related positions. These results suggest that AI is changing the composition of tasks and skills within exposed establishments. Still, the effects on the labour market are not yet evident, possibly due to the early stage of AI adoption. This research highlights the need for further studies to understand the impacts at the establishment level and the lack of aggregate effects and explore other potential implications of AI.

Gmyrek et al. (2023) stood out for their methodological difference, as they replaced the assessment of the level of automation of a task, previously carried out by experts, with the use of a Large Language Model (LLM), namely the GPT-4 model. This replacement was motivated by suspicions that experts tended to overestimate how automatable a specific task might or might not be (Karger et al., 2023), and by the "surprising closeness of the GPT4 predictions to the judgements made by a group of AI experts" (Gmyrek, Berg, & Bescond, 2023, p. 10), showing a correlation between the two (Eloundou et al., 2023). Although this study focuses on the automation of tasks only in the adoption and integration of generative AI, it is nonetheless pertinent given the similar nature of the final objectives of its work. Therefore, to adopt a methodology similar to the previous ones and considering the previously mentioned aspects, they used the GPT-4 model to determine the risk of automating the numerous tasks in the labour market.

The results obtained confirmed the change in the automation landscape previously highlighted by Lassébie and Quintini (2022), reinforcing that the emergence and evolution of models such as the Generative Pre-training Transformer (GPT) will begin to affect more and more jobs, particularly the most highly qualified occupations. As a result, groups of workers previously considered untouchable regarding the risk of some tasks disappearing are now facing this possibility. The study highlighted that Professionals, encompassing many tasks traditionally considered part of "knowledge work," are now at risk. This includes tasks related to drafting documents, analysing text, and searching for information, which was previously thought to require human cognitive abilities but can now be partially automated by advanced AI technologies. Similarly, service and sales workers face increased exposure to automation, with some tasks surpassing the medium exposure threshold. Tasks such as making and confirming reservations, responding to inquiries, and providing various customer service activities are now increasingly susceptible to automation. Furthermore, even elementary occupations, traditionally seen as safe from automation due to their manual nature, are now at risk. Tasks like maintaining records, preparing invoices, and performing other administrative duties can be automated, reducing the need for human intervention in these roles. These findings underscore the need for proactive policies and training programs to support workers in transitioning to new roles and acquiring skills that are less susceptible to automation.

Xu et al. (2023) presented a new angle on calculating the risk of automation at a task level by fine-tuning a pre-trained, high-quality machine learning model for NLP called Bidirectional Encoder Representations for Transformers (BERT). The training of this model was based on the classifications of various experts using a specific scale, allowing the model to learn and replicate the experts' criteria when classifying each task. After refining the model to achieve the best possible performance in approximating expert classifications, O*NET data was used, as in previous studies, to obtain a view of the level of occupations that group different tasks. The results showed that approximately 25% of the jobs present in O*NET present a significant risk of automation.

Performance metrics further highlighted the model's robustness. The BERT model achieved approximately a precision of 80%, recall of 80%, and F1-score of 79% when both datasets (O*NET, ESCO) were combined. This indicates the model's robustness across diverse and large-scale datasets. When compared to other models, BERT outperformed traditional machine learning models like Logistic Regression (precision: 59%, recall: 59%, F1-score: 59%) and neural network models like GRU (precision: 76%, recall: 76%, F1-score: 76%), underscoring its superior performance in their classification task. Since the automation of occupations is related to the description's semantic and contextual understanding of their nature (Frank et al., 2019, as mentioned in Xu et al., 2023) and since BERT is a high-quality model for this type of task (Devlin et al., 2018), the authors proved its good performance in understanding the job descriptions present in O*NET and ESCO.

2.2.3. Data Source Considerations: O*NET vs. ESCO for Portugal

As we have seen so far, studies aimed at OECD member countries, of which approximately 60% are located in Europe (OECD, 2024), use data from the O*NET (Arntz et al., 2016; Nedelkoska and Quintini, 2018; Lassébie and Quintini, 2022). Although the applicability of this data to the European reality has been proven, showing a fairly pronounced similarity (Handel, 2012) and existing a crosswalk created by the entities themselves that allow equivalences to be established between occupational categories, skills and tasks, the O*NET data source was originally for the reality of the United States of America.

Even in studies aimed explicitly at Portugal, the choice falls on O*NET rather than ESCO, as was the case with Fonseca et al. (2018), who studied the polarisation of the labour market in Portugal. After analysing the possible reasons for this preference, we concluded that this preference for O*NET is due to the greater detail of the data presented, including more metrics and indicators helpful to users worldwide, such as the importance and frequency of a task for a respective occupation.

For this reason, with the help of the model with high semantic and contextual capabilities, such as BERT (Xu et al., 2023), to create the missing variables based on the data present in O*NET, we have decided to use ESCO data as the main data source in our project, as we will discuss in more detail in methodology chapter (Section 3). In this way, we will take advantage of the structure of occupations aimed at the European labour market, where Portugal is located, while also providing a solution to some information gaps by making new indicators available for the existing ESCO database.

2.3. Recommendation Systems

Integrating recommendation systems has become essential in the workplace, helping with career transitions, skills alignment, and job satisfaction. We highlight three notable contributions, each proposing significant methodologies for dealing with the complexities inherent in job transitions and recommendations for skills and competences.

Dave et al. (2018) presented a recommendation system to deal with the complexities of job transitions and skills requirements. Their approach aims to reduce skills gaps and facilitate informed job transitions in increasingly dynamic industries by identifying relevant skills for specific jobs and calculating similarities between skills to suggest relevant skill acquisitions in a given job transition. Dividing it into three key graphs (job transition, job-skill, and skill co-occurrence), they proposed a joint representation learning framework to create shared vector spaces for jobs and skills. This framework allowed for precise recommendations of jobs and skills, considering the complex relationships between jobs and skills. The system's strengths lie in its ability to unify multiple sources of information, provide representation vectors, and surpass existing methods in job and skills recommendations based on evaluating professional summaries and job offers from the CareerBuilder platform. However, a limitation of the system recognised by the authors is its static nature, since the model learns the representation vectors based on the existing entities in the input graphs, needing to be retrained every time new jobs and skills are inserted, which represents a challenge for real-time adaptability to the evolution of labour markets. They suggest implementing inductive learning in the system to deal with these updates in any labour market as a future solution.

Just as Dave et al. (2018) used data from job portals, Tavakoli et al. (2020) took the same approach. They developed an Open Education Resources (OER) recommendation system to help individuals acquire the skills required for a particular job, facilitating a possible transition in the labour market. Firstly, the system applies classification and text extraction techniques to deconstruct job vacancies into core skills components, identifying the focus for learners along their learning journey. Secondly, it uses a hybrid OER recommendation system, tailoring personalised learning content to guide learners towards progressing towards these skills objectives. The initial evaluation of the prototype was centred on the professional areas of data science and mechanical engineering, which generated more than 150 recommendations, with around 77% of them considered valuable by experts in the field. The fact that this system has been integrated into a dashboard has allowed the experience to differ for each user, allowing each learner to search for their current or desired job, display the list of skills required and even define their level of specialisation for each skill.

Dawson et al. (2021) developed an innovative recommendation system to facilitate professional transitions during socio-economic changes, motivated by crises such as the COVID-19 pandemic and substantial structural changes at work caused by emerging technologies such as Artificial Intelligence. They developed a new methodology called Skills Space, which quantifies the similarity between occupations based on their skill sets extracted from an extensive dataset of Australian job advertisements from 2012 to 2020. Their professional transition recommendation system uses Skills Space measurements to predict professional transitions with an accuracy rate of 76%, taking into account the asymmetries inherent in professional movements, providing a guide to skills and competences, based on the distance and importance between the jobs involved in the transition, making it safer and more efficient.

Part of our final recommendations were inspired by these works, especially by the work developed by Dawson et al. (2021), as we will explain in more detail in the modelling (Section 3.3.2) and results (Section 4.4) sections. However, it is crucial to recognise the limitations of their approach, including potential biases arising from the relatively small sample size of the dataset, the aggregation of results at a specific occupational level, which may overlook differences in more particular occupations, and the other factors that can influence a transition in the labour market, such as an individual's work experience, education or even job satisfaction.

2.4. Summary

To encapsulate the key insights and frameworks discussed in this literature review, this summary will highlight the foundational elements that underpin the research and their implications for the project.

Understanding the definitions and structures of O*NET and ESCO is crucial for the analysis carried out in this project. The data provided by these systems allows for a consistent and comparable examination of occupational information across different sectors and geographic realities. By leveraging these definitions and the crosswalk between them, the research gains access to the structures of the various occupations and constituent tasks, as well as identifying how they differ. For Career Xplorer, these definitions and the crosswalk support the accurate prediction of automation risks and the development of personalised career transition guides, helping Generation X professionals in Portugal navigate the evolving technological landscape. That said, O*NET and ESCO are essential tools for labour market analysis, in which their definitions and structures, combined with the intersection that links them and projects such as ours, offer valuable information to various stakeholders, including policymakers, employers and educators.

In addition, this literature review identifies three different methods for measuring automation progress in the workplace. Some are based on the knowledge of experts (Frey and Osborne, 2013; Arntz et al., 2016; Lassébie and Quintini, 2022; Xu et al., 2023), while others replace these experts by using the GPT-4 model (Gmyrek et al., 2023), once the similarity between the two has been proven (Eloundou et al., 2023). Finally, some authors use data from numerous online job offers to create an indicator that represents the demand over time for AI-related skills and competences, thus determining the level of automation of tasks and, consequently, the occupations that encompass them (Dawson et al., 2021; Acemoglu et al., 2022). Despite the different methods, there is an increase in the percentage of jobs at risk globally, which highlights the importance of carrying out and analysing studies and projects of this kind, which analyse the current situation and develop recommendation systems for individuals at risk.

As illustrated in Figure 2.1 and previously mentioned, the studies have reported different percentages of jobs at high risk of automation, overall representing an increase over time. Notably, Frey and Osborne (2013) reported a significantly higher estimate of 47%, which has been proven by subsequent studies to be an overestimation, mainly due to their less adequate Occupation-Based Approach and an exaggerated technological optimism. In contrast, Task-Based Approaches, such as those used by Xu et al. (2023) and Lassébie & Quintini (2022), provide more accurate and lower estimates of 25% and 28%, respectively. These variations, particularly the increase in high-risk occupations, underscore the growing presence of automation and the need for ongoing, precise assessments to better prepare for its impacts.

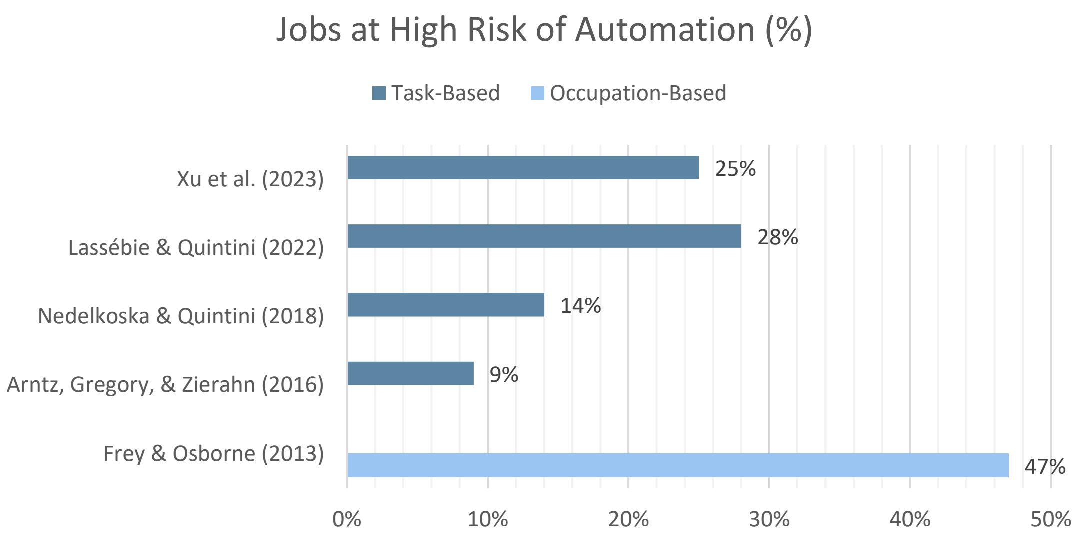

Figure 2.1 – Comparison of High-Risk Automation Job Estimates by Study and Approach

Concerning the interactive recommendation systems, all the systems analysed served as valuable resources for policymakers, educators, and job seekers to navigate career transitions efficiently. Although Tavakoli et al. (2020) integrated an interactive dimension through a dashboard, many recommendation systems do not incorporate this feature, resulting in static and non-personalised systems that are less appealing and more expensive to update. In short, the advantages of interactive recommendation systems are manifold. They allow personalisation through user interaction, enabling active involvement and feedback to improve subsequent recommendations. This iterative process ensures the continuous evolution of the system, aligning it with evolving user preferences and mitigating information overload. Additionally, these systems offer real-time adaptability, dynamically adjusting to the user's needs and preferences and offering truly personalised suggestions.

That said, it was crucial for this project to incorporate this important interactive dimension. We went beyond the level of interactivity presented by Tavakoli et al. (2020), who developed a dashboard, by opting to develop a web application. This approach not only enhances the user experience but also allows for the integration of a wider range of features, increasing the system's flexibility and potential for future updates.

To summarise, by integrating these comprehensive systems and methodologies, our project not only aims to upgrade the ESCO database to match the level of detail of O*NET but also to develop an innovative interactive platform. This platform provides robust integrated career guidance, specifically tailored to the needs of Generation X professionals in Portugal, ensuring that they are well-prepared to face the challenges and opportunities presented by the evolving labour market.

As we transition to Section 3, the focus shifts from theoretical foundations to practical implementation. The next section describes the methods used to build the recommendation system, elucidating the steps and techniques that translate conceptual structures into operational systems.

3. METHODOLOGY

The Cross-Industry Standard Process for Data Mining (CRISP-DM), first conceptualised in 1996 and officially unveiled in 2000 by Chapman et al., continues to serve as the fundamental framework for data analytics initiatives. Its structured approach unfolds across six iterative phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, each critical to the lifecycle of a data mining project. The framework's applicability extends to diverse applications, including recommendation systems, offering a methodical pathway through the analytics process. This adaptability is particularly pertinent to my research on AI's impact on Generation X careers in Portugal, providing a comprehensive blueprint that aligns with the exploratory nature and specific objectives of my study (Chapman et al., 2000; Martínez-Plumed et al., 2019). To frame the implementation of CRISP-DM within this research, an illustrative diagram is presented, which delineates a high-level overview of the CRISP-DM phases, carefully tailored to the scope of this project. The diagram serves as a visual guide, succinctly mapping out the key components and data sources that will be employed while showcasing the logical progression from the initial phase of business understanding through to the final stage of deployment.

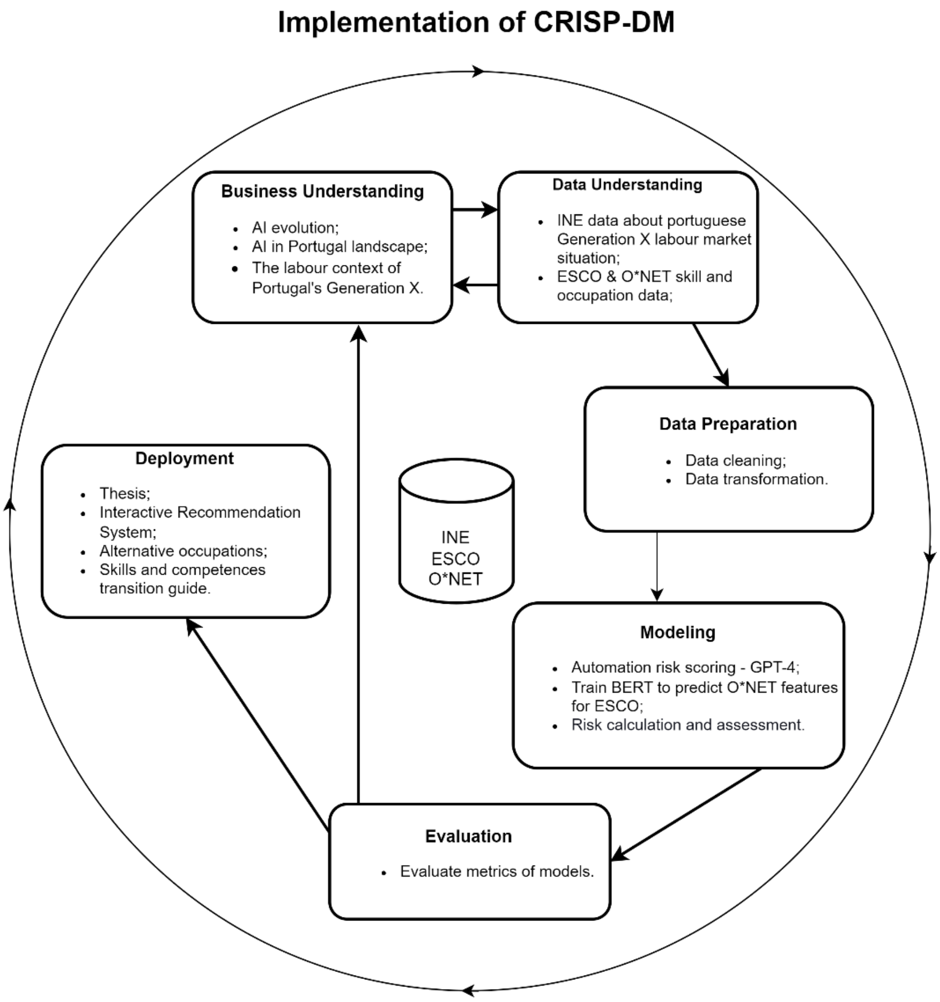

Figure 3.1 – Implementation of CRISP-DM

3.1. Business Understanding

In the Business Understanding phase, in addition to examining specific challenges unique to Generation X individuals, such as the increasing divorce rate, we analyse the distribution of Generation X within the Portuguese labour market, utilising data from INE (National Statistics Institute). This phase is crucial for defining the context and scope of the research, ensuring the project addresses the most prevalent realities faced by Generation X professionals in Portugal. We aim to encompass as many professionals as possible while avoiding data overkill, which would not only complicate the analysis but also make hyperparameter tuning during the model training phase and the use of the GPT-4 API extremely expensive. Therefore, we are focusing on the three most representative groups of this generation to demonstrate how our methodology and our prototype web application can effectively support the entire process of a career transition (Figure 3.2).

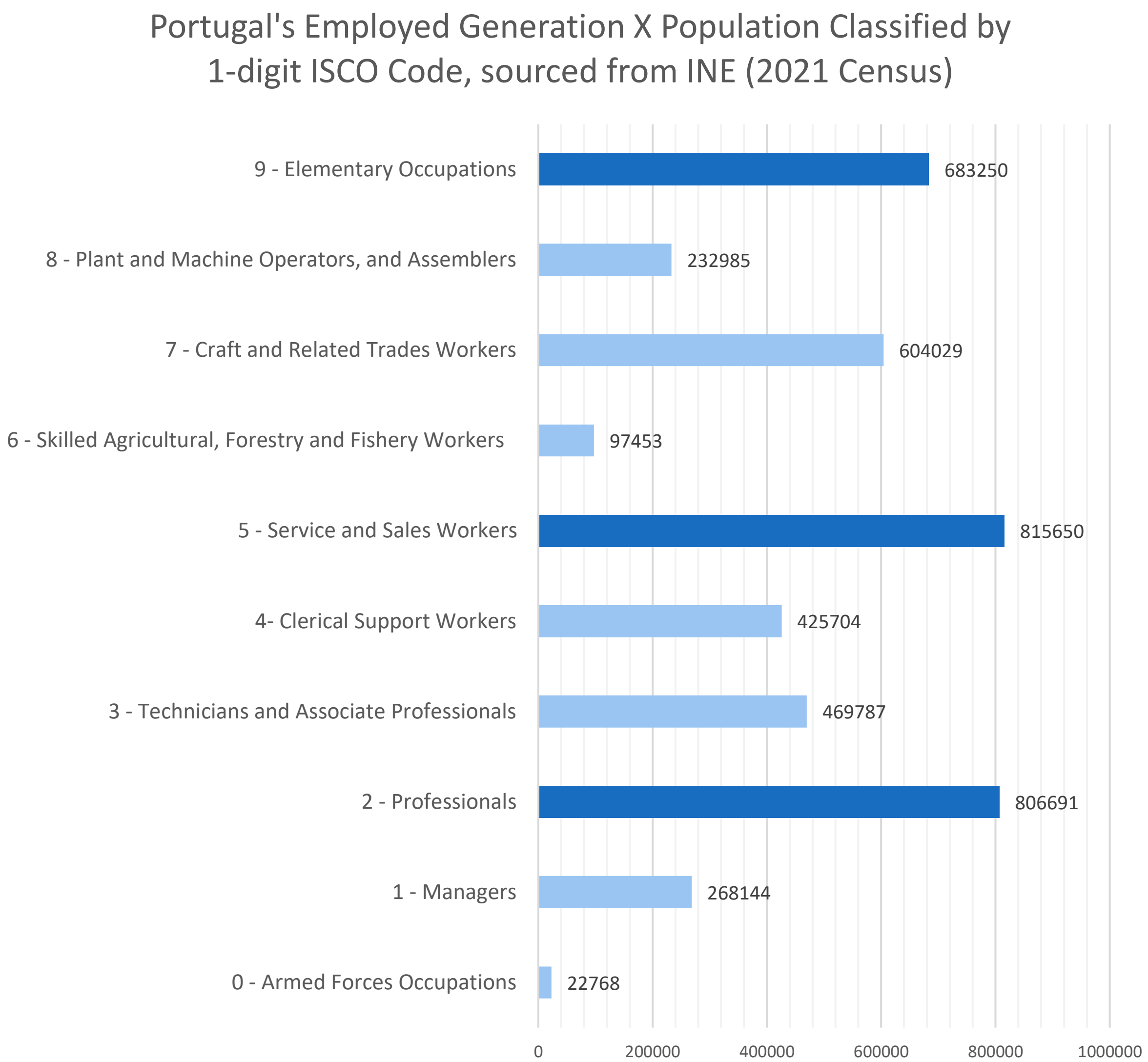

Figure 3.2 – Portugal's Employed Generation X Population Classified by 1-digit ISCO Code, sourced from INE (2021 Census)

3.2. Data Understanding & Preparation

Before analysing the structure of the data and the treatment it has undergone, we present the diagram below, which offers a concise and simplified representation of the data landscape, outlining the sources from which we gathered information and the specific objectives associated with each dataset. By presenting this visual overview, we aim to give readers a clear understanding of the fundamental data elements that guide my research, setting the stage for a more in-depth exploration of the methodology.

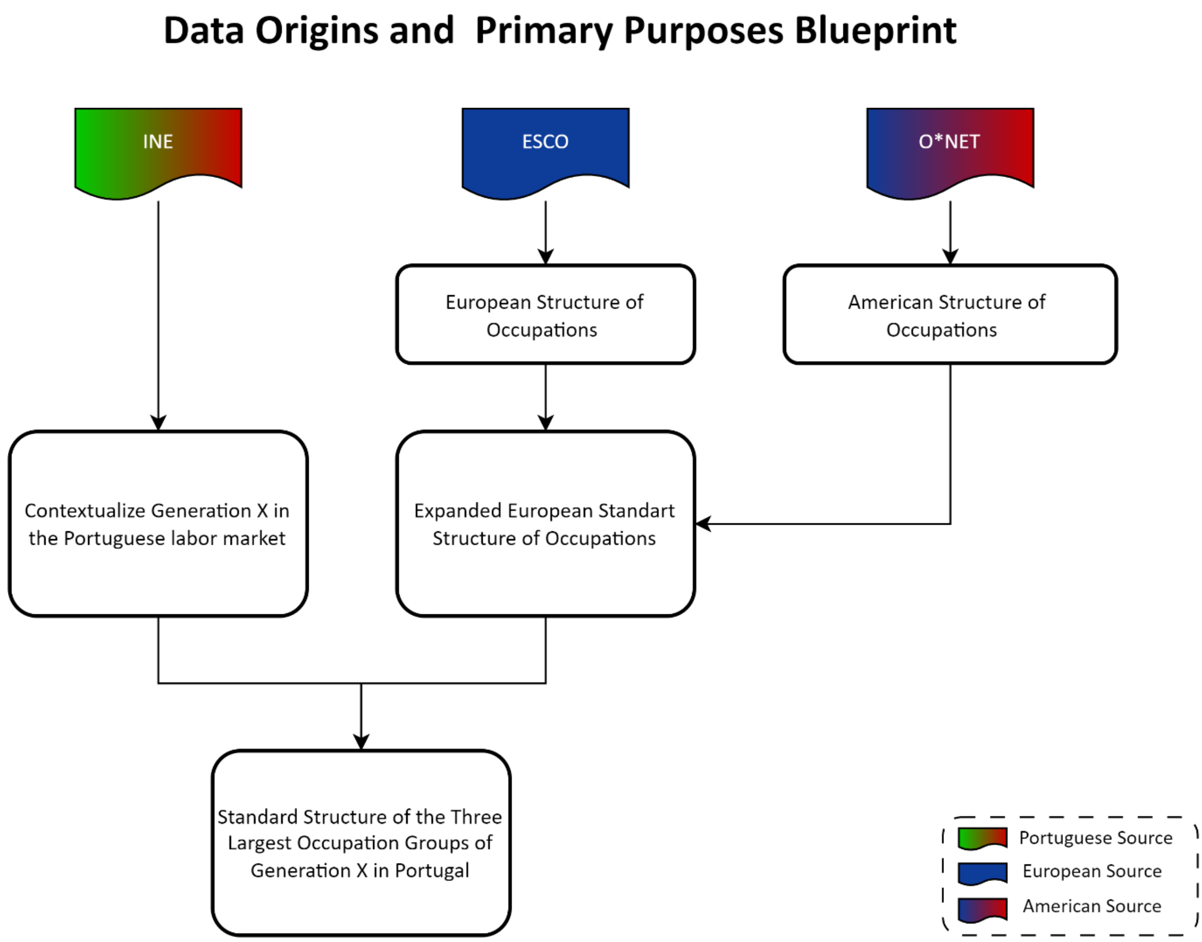

Figure 3.3 – Data Origins and Intentions Blueprint

As already mentioned, the choice of data source considered Portugal's geographical context. Instead of adopting the data on the structure of occupations developed for the USA, which offers a higher level of detail compared to the data originally associated with the European reality, we decided to use this more detailed American source (O*NET) only as a complement in the process of filling in the gaps in the data from the European source (ESCO), as illustrated in Figure 3.2. Despite the compatibility of American information with the European reality, supported by similar studies, we decided to use the structure of occupations created directly for the European reality and resolve the main reason that other studies choose one data source over the other, namely the level of detail between them. In this way, we provide a more complete structure of European occupations, encouraging its use in future studies related to the European labour structure, rather than using the American one for the European reality, which also promotes a closer approximation to reality in the results obtained.

3.2.1. Crosswalk between O*NET and ESCO

As discussed in previous chapters, key term definitions can slightly differ between the O*NET and ESCO, particularly regarding occupations. To ensure consistency and coherence in Career Xplorer, it was necessary to harmonise information. This was achieved by integrating O*NET and ESCO data through a crosswalk that links the two systems.

The crosswalk provides several advantages. Firstly, mitigates domain shift by incorporating ESCO-specific language and terminologies into training data, which would otherwise occur if only O*NET data were used. This alignment enhances the model's ability to generalise and perform accurately when applied to ESCO data. Secondly, allows the use of detailed task descriptions, importance, and frequency ratings from O*NET while contextualising them within ESCO occupations. This combination results in a richer and more informative training dataset, improving the model's performance and applicability within the ESCO framework. This connection between O*NET and ESCO can be verified in Appendix A, Figure A1.

Table 3.1 compares occupational data between the two data sources, highlighting the number of unique occupations, mapped occupations, and filtered data for Generation X in Portugal. This crosswalk establishes a direct link between them, allowing for efficient and accurate mapping. Its use was crucial for building the final dataset used in the training phase of fine-tuning the models and is essential for justifying the predictions.

	O*NET	ESCO
Classification version	O*NET-SOC 2019	ESCO V1.1.0
Total unique occupations	1016	3008
Unique occupations mapped	958	2997
Top 3 Gen X occupation groups in Portugal (filtered)	709	1262
4-digit occupation	399	134

Table 3.1 – Comparison of O*NET and ESCO Classification Systems

3.2.2. ESCO Database

Regarding the ESCO database, various datasets relating to different levels of detail of the occupations were analysed. The *ISCOGroups* dataset provides information on the categorisation of occupations, ranging from the most comprehensive levels (1-digit ISCO code) to the most specific (4-digit ISCO code), and a description of these groups. From the descriptions at the most specific group level (4-digit ISCO code), we extracted the tasks associated with each occupation group using data preprocessing techniques, similar to previous studies that have broken down job offers into key skills components (Dave et al., 2018; Tavakoli et al., 2020; Dawson et al., 2021). Appendix A, Figure A2, illustrates the result of this process.

The *Skills* dataset should also be highlighted, as it was used in the development of the skills and competences guide, which will be discussed in more detail later. This dataset includes all the skills and competences and their respective descriptions. It was cross-referenced with the *SkillsHierarchy* dataset, which organises the skills and competences at different levels. In the case of skills and competences, we focused our analysis on the individual skills and competences, which are at a more specific level than the group level. This approach contrasts with the method used to calculate the risk of automation for different occupations, where the focus was on the tasks within the 4-digit ISCO groups rather than individual occupations. This distinction arises because the differences between some occupations, such as animal chiropractor and animal osteopath (both belonging to the 4-digit ISCO Veterinarians group), are not significant, unlike the differences in skills and competences. Therefore, given the homogeneous nature of tasks within occupations in the same 4-digit ISCO group, we chose to identify tasks at this group level rather than individual occupations to calculate the risk of automation at the task level, thus avoiding unnecessary redundancy.

Finally, a cross-referencing was carried out with the occupation's datasets, to understand which skills and competences were associated with which occupations. As can be seen in Appendix A, Figure A3, this process resulted in a consolidated dataset, facilitating the subsequent analysis of the relationships between skills, competences, and occupations.

3.2.3. O*NET Database

The data analysis by O*NET primarily focused on the *Task Ratings* dataset, which maps O*NET-SOC (Standard Occupational Classification) codes to the ratings of tasks associated with each occupation. These ratings are available on different scales, such as the relevance, importance and frequency that each task has within a specific occupation.

Initially, we worked with a set of 873 occupations and 17983 tasks. After cleaning the data, which included the removal of unreliable data, filtering only relevant values, and cross-referencing data with the O*NET to ESCO crosswalk, we reduced the set to 243 occupations and 5137 tasks. Subsequently, the *Task Ratings* dataset was divided into two different datasets according to the different scales used for our analysis: frequency and importance.

As we can see in Table 3.2, originally the frequency scale was defined in seven different levels, ranging from "Yearly or less" to "Hourly or more". However, to simplify the analysis and interpretation for users and to facilitate predictions made by the models, this scale was modified to three levels: "Daily or more frequently", "More than monthly to weekly", and "Yearly or more".

Original Frequency Description	Merged Frequency Descriptions
Yearly or less	Yearly or more
More than yearly	
More than monthly	More than monthly to weekly
More than weekly	
Daily	Daily or more frequently
Several times daily	
Hourly or more	

Table 3.2 – Frequency Description Merging

As shown in Table 3.3, the importance scale, originally defined on a scale from 1 to 5, was transformed into a scale of 3 to 5 after the data cross-referencing, cleaning and pre-processing stages. The redefined scale categorises tasks as “slightly important”, “important”, and “very important” disregarding the “not important” categories due to the nature of the data. This gap in the scale turns out to be irrelevant, as the main objective of this process is to present a standard occupational structure in which users can customise the level of importance using the full scale (from 1 to 5) in the web application, depending on their situation, as we will demonstrate further in Section 4.6.

To address the imbalance in the dataset and ensure a better performance of the BERT model, we increased the number of instances in each importance category. This was achieved through a paraphrasing process, utilising GPT-4 to rephrase task descriptions with different wording while preserving their original meaning (Xu et al., 2023). This augmentation process ensured that each importance category had sufficient representation, enhancing the robustness of our model performance, as we will see in more detail in Section 4.1.

Importance Scale	Original Data	O*NET Processed Data	Augmented Data	ESCO + O*NET Refined Data
1	2	-	-	-
2	146	-	-	-
3	2716	853	5160 (+4307)	4459
4	12656	3506	4022 (+516)	3427
5	2716	736	3996 (+3260)	3771

Table 3.3 – Importance Scale Data Distribution Across Different Processing Stages

These two datasets were subsequently cross-referenced with the *Occupations* dataset using the occupation code feature (O*NET-SOC Code). This process enabled the creation of a more robust and relevant dataset, now enriched with a detailed description of each occupation. The results of this integration can be seen in Appendix A, Figures A4 and A5, in the case of the frequency dataset.

3.2.4. Training Dataset

Given our specific goal of fine-tuning a pre-trained BERT model to predict the importance and frequency of tasks within an occupation, we decided to utilise a combined dataset for the training phase. Initially, we considered using only O*NET data to train the model and then applying it to ESCO data. However, by leveraging the crosswalk that directly links occupations between O*NET and ESCO, we adopted a different approach.

By using ESCO occupation names and descriptions combined with O*NET task descriptions for model training, we aimed to better align the model with the ESCO framework. This approach offers several advantages over the alternative method of using only O*NET data for both training and application. Firstly, training the model with ESCO occupation names and descriptions ensures consistency and relevance to the target domain. Since our goal is to predict task importance within ESCO, using ESCO-specific language and terminologies during training reduces the domain shift that might occur if only O*NET data were used, due to the different geographic context of these two sources. This alignment improves the model's ability to generalise and perform accurately when applied to ESCO data. Secondly, the crosswalk between O*NET and ESCO provides a valuable link, enabling us to take benefit of the detailed task descriptions, importance, and frequency ratings from O*NET while contextualising them within ESCO occupations. This combination allows us to benefit from the richness of O*NET's task data and the specificity of ESCO's occupational framework, resulting in a more robust and informative training dataset.

In contrast, the alternative approach of using only O*NET data for training has its limitations. Although O*NET offers comprehensive occupation and task descriptions, the potential differences in language and focus between O*NET and ESCO have led to challenges in model generalisation. As we will discuss in Section 4.1.1, the model struggled more to accurately interpret and predict the importance of tasks within the ESCO framework due to these discrepancies.

Furthermore, using a single data source (O*NET) may not fully capture the unique characteristics and nuances present in ESCO data. As we saw previously, ESCO and O*NET have slight differences in the terms and concepts used, such as occupations, and training solely on O*NET data could result in a model that is less aligned with the specifics of ESCO occupations. In addition, we were fine-tuning the model to predict these indicators (importance and frequency) of tasks for different occupation levels, specifically at the 4-digit ISCO code level in ESCO. If we trained with the O*NET occupation structure, we would be using the most detailed level, the occupations themselves, which did not align perfectly with the ESCO framework used.

As a result of our chosen approach, the 243 O*NET occupations were mapped to 108 equivalent 4-digit ISCO code ESCO occupations. This mapping provided a coherent and aligned dataset for training, as shown in Appendix A, Figure A6, ultimately enhancing the model's performance and applicability within the ESCO framework.

In summary, by integrating ESCO occupation information with O*NET task descriptions, we have developed a more targeted and effective methodology for fine-tuning our pre-trained BERT model, ensuring that it meets the specific requirements and nuances of the ESCO data.

3.3. Modeling

During the modeling phase, our methodology leverages the latest advancements in artificial intelligence, specifically large language models, which have been rapidly evolving and expanding their applicability. This phase is divided into two main chapters. In the first chapter, we focus on calculating the automation risk of different occupations by analysing their constituent tasks. In the second chapter, we discuss the recommendation of alternative occupations and provide a skills and competences guide to support users in transitioning from one occupation to another. This recommendation considers the previously calculated automation risk and the similarity of skills and competences. For a better understanding of this process, a more detailed view is presented in Figure 3.4.

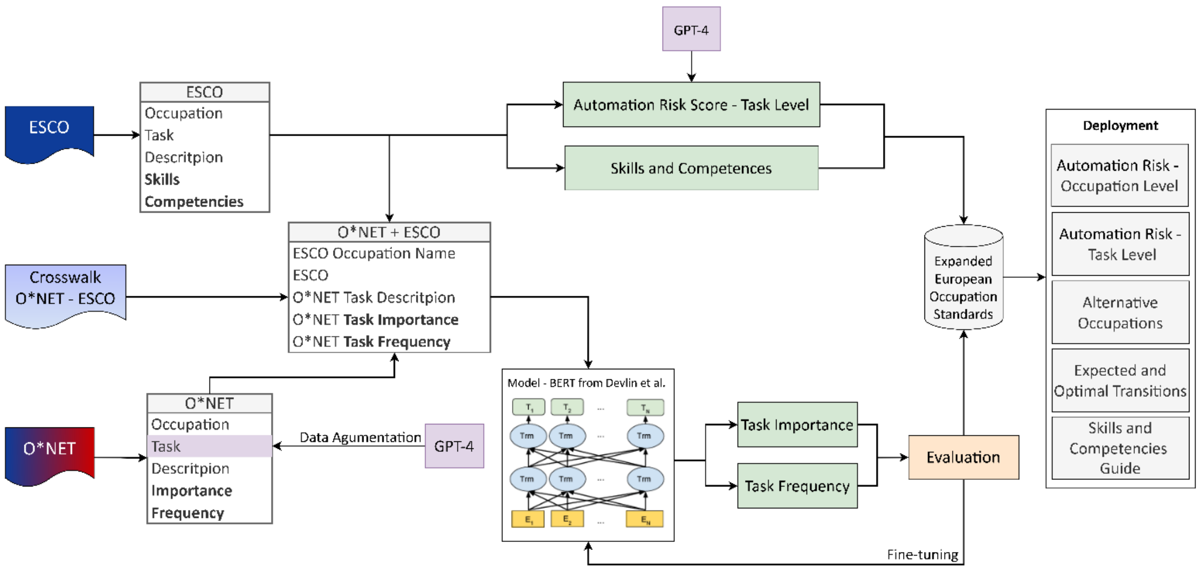

Figure 3.4 – Workflow of the Automation Risk Assessment and Career Transition Recommendation System

3.3.1. Automation Risk – Occupation Level

For the calculation of the automation risk of a particular occupation at the task level, we adapted a formula used by Lassébie & Quintini, 2022. This adaptation comprises three components: the automation risk of a task, the importance of the task to the occupation, and the frequency with which the task is performed in that occupation. Therefore, the automation risk score of each occupation o=1, ..., O is then calculated as:

$$automation\,risk\,score = \sum_{i=1}^{I} automatability_i \times \text{importance}_{\,io} \times frequency_{\,io} \qquad (3.1)$$

Where *i* indicates a specific task, *automatability i* is the automation risk score of a task, *importance io* is the importance value of task *i* in occupation *o*, and *frequency io* is the frequency with which task *i* is performed in occupation *o*.

In the following subchapters, we will go into more detail about each element of the formula used to calculate the automation risk score of different occupations. First, we will explore the concept of automation, which assesses the risk of a task being automated. Next, we will examine the importance and frequency of each task within the respective occupation, determining how crucial a task is and how regular it is in each occupation. Together, these components form the automation risk score calculation at the task level for each occupation.

3.3.1.1. Automation Risk – Task Level

As we can see from the formula presented above, our principle defines the risk of automation of a task as an indicator that is independent of the occupation in which the task is carried out, meaning that whatever occupation the task is carried out in, the risk of automation is the same.

As we have seen previously, the technological advancements achieved to date have made it feasible to use LLMs to measure the level of automation of certain tasks (Eloundou et al., 2023; Gmyrek et al., 2023). Therefore, due to the inaccessibility of expert assessment, we opted to use a multimodal language model developed by OpenAI, named GPT-4. This process involved using the API (Application Programming Interface) to request that GPT-4 generate a score between 0 and 1 for each task, representing the risk of automation that a task presents based on the task's description, the occupation it belongs to and its 4-digit ISCO code (adapted from Gmyrek et al., 2023). The request was described as follows:

> *{"role": "system", "content": "You are a highly specialised AI trained to evaluate the potential for automation of tasks within occupations."},*
>
> *{"role": "user", "content": f"Considering the job task: '{task}', which is related to the occupation: '{occupation_name}' and has the ISCO code: '{isco_code}', provide a numerical score between 0 and 1, inclusive, representing the potential for automation with GPT and artificial intelligence technology. Provide only the numerical score without any additional explanation."}*

To ensure the reliability of our automation risk scores, we tested the consistency of the GPT-4 model's predictions. We randomly selected five tasks and ran the scoring process 20 times for each task. We then calculated the mean score and SD (standard deviation) for each task to measure consistency. As shown in Table 3.4, the scores were very consistent, with SDs not exceeding 0.10, meaning that in real-world application scenarios, our error margin for each task's automation risk is not greater than 10%, ensuring a high level of reliability in our predictions. For instance, the task "filling storage areas of vending machines and collecting money from their containers" had a mean score of 0.88 and an SD of 0.04, indicating stable scores across multiple runs. This consistency suggests that GPT-4 provides reliable automation risk assessments based on task descriptions and occupational contexts.

4-Digit Occupation Code	Task	Mean ± SD
9623	Filling storage areas of vending machines and collecting money from their containers	0.88 ± 0.04
5151	Making beds, cleaning bathrooms, supplying towels, soap and related items	0.84 ± 0.04
9213	Digging and shovelling to clear ditches or for other purposes	0.86± 0.03
5329	Lifting, turning and moving patients and transporting them in wheelchairs or on moveable beds	0.21 ± 0.09
9311	Assembling and dismantling mining equipment	0.76 ± 0.09

Table 3.4 – Test of GPT-4 Score Consistency (20 task-level predictions)

We choose tasks for automation risk calculation instead of skills and abilities because tasks provide a detailed view of the activities within an occupation. Evaluating the automatability of each task allows for a precise assessment of which job components are most vulnerable to automation. This granularity ensures that the risk assessment is accurate and actionable, identifying the exact functions that may be replaced by AI or machines. While skills and abilities offer a broader perspective on an individual's competencies, they lack the specificity needed for an accurate risk assessment. Skills and abilities are more abstract and transferable across jobs, making it harder to directly assess their susceptibility to automation. This could lead to a less detailed understanding of automation risk, making it challenging to identify and address specific job components at risk. Therefore, tasks are preferred for a more accurate and actionable automation risk calculation.

3.3.1.2. Importance and Frequency of a Task

To determine the importance and frequency of a task in a specific occupation, we were inspired by the approach taken by Xu et al. (2023). They fine-tuned a pre-trained BERT model, leveraging its deep learning capabilities, and trained it with data from O*NET and ESCO to learn the criteria experts use to classify the risk of task automation.

As illustrated in Figure 3.5, while adapting Devlin et al. (2019) single-sentence classification task, we also fine-tuned a pre-trained BERT model, but for a more complex assignment, such as predicting the importance and frequency components of a task in the ESCO database, based on the corresponding task in the O*NET database. The input variables for both models consisted of three different features: occupation name, occupation description, and task description.

To provide context and help the model interpret the input information while reducing complexity, we combined the three features into a single text field, separating each feature with a personalised token - [OCC_NAME], [OCC_DESC], and [TASK_DESC] for occupation name, occupation description, and task description, respectively. This combined field, the personalised tokens, and the standard BERT tokens ([CLS]; [SEP]), enabled BERT to more easily identify the differences between these elements and better understand their relationships, as reflected in the results presented in Section 4.

Among the several variations of BERT available, we opted for the BERT-base-uncased. This choice was not only sufficient but also effective for our needs, demonstrating excellent results. A more complex variation was unnecessary given the characteristics of our data and our objectives. For instance, the case of the words was not important, allowing the model to treat 'Task' and 'task' as identical, which simplifies processing and reduces complexity.

In our implementation, the result representing the class label is described as the importance and frequency of a task within an occupation, as illustrated in Figure 3.5.

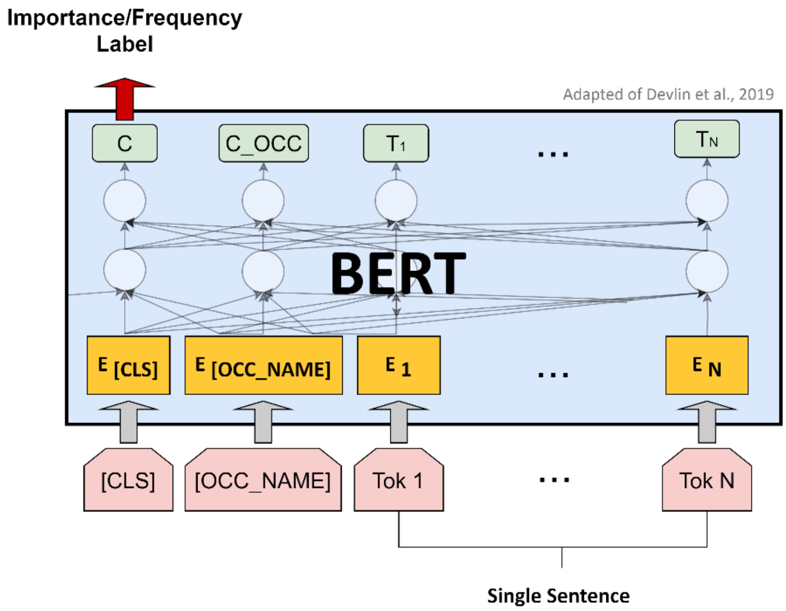

Figure 3.5 – Fine-Tuning BERT on Single Sentence Classification Task. Adaptation of BERT for predicting task importance and frequency within occupations by combining occupation name, description, and task description into a single input field with personalised tokens.

Regarding the architecture of our model, it includes one main layer to capture the complex relationships in the data, with dense, normalisation, dropout layers, and an activation function. The output layer used a softmax activation function to ensure the attention scores were normalised. This architecture allows the model to assign higher attention scores to more important words, enhancing the model's ability to focus on critical information within the text (Xu et al., 2023).

As our model architecture and the detailed results will show, we focused extensively on reducing complexity and other factors that could contribute to overfitting and poor performance. Achieving good results was crucial, as we are applying this model to unseen data with the objective of improving the ESCO database in a reliable and high-quality manner, benefiting not only our work but also future research in this field. Therefore, to avoid overfitting, we split the dataset into training and validation sets while reserving a percentage for a test set, ensuring we had a portion of data to evaluate the model's performance on unseen data. For the same reason, we employed a Stratified K-fold cross-validation approach to ensure a more robust evaluation. This method helps evenly distribute the different importance and frequency levels across folds, providing a more consistent performance estimate.

To further improve model performance, we utilised several techniques, including custom early stopping, learning rate reduction on plateau validation loss and accuracy, and model checkpointing. Additionally, we incorporated callbacks to monitor various evaluation metrics, providing a comprehensive evaluation of the model's performance, as we will discuss in Section 3.4.

By leveraging this robust approach, we addressed the gaps in the ESCO database, resulting in a more complete dataset that is comparable to O*NET. This enhancement promotes greater independence from O*NET data for research of this nature and, consequently, encourages greater use of ESCO data in European contexts.

3.3.2. Similarity between Occupations and Skills and Competences

As Dawson et al. (2021) state occupations are “similar when their most important skills are similar” (p. 4). Based on this insight, we developed our recommendation system, which suggests the best alternative occupations for users based on the similarity of skills and competences, as well as automation risk scores. This system also provides a transition guide with skills and competences ranked by relevance within the occupations and distance from each other. The aim was to design a more efficient and effective path for a smooth and enjoyable occupation transition, thereby boosting user confidence throughout the process. A better occupation recommendation has high skill[1] similarity between occupations and a lower automation risk score.

This analysis integrates quantitative data transformation techniques and clustering algorithms to reveal patterns and relationships among different occupations. The methodology follows a

[1] For simplicity, “skill” will be used throughout this paper to refer to both a single skill or competence and the plural form encompassing both skills and competences.

structured approach, leveraging the relevance of skills to occupations, dimensionality reduction and clustering processes.

Each skill has two levels of relevance within an occupation: essential and optional. This qualitative data was transformed into numerical values, assigning a score of 1 to “essential” skills and 0.5 to “optional” ones. This transformation allowed us to create a matrix where rows represented occupations and columns represented skills, with the cells indicating their relevance to the corresponding occupation. Missing values in this matrix were filled with zeros, representing the absence of a skill, thus implying it is not associated with the occupation. Unlike tasks, which are highly specific to their respective occupations and may appear different due to details such as their fields of application, skills are more transferable between occupations. This commonality enables the calculation of distances not only between skills themselves but also between different occupations, resulting in comprehensive data for analysis.

To visualise and analyse the similarities between skills, we initially used Euclidean distance to calculate the pairwise distances. Euclidean distance measures straight-line distances in a multi-dimensional space, which can be effective in some contexts. However, we found that Euclidean distance often produced misleading results, particularly in cases where occupations had vastly different numbers of skills. For instance, an occupation with many skills could appear more similar to an occupation with few skills simply by sharing a few common ones, which doesn’t represent reality because it ignores the proportionality and overall distribution of skills across occupations.

Recognising this limitation, we shifted to using the Weighted Jaccard Similarity metric. This metric measures the similarity between two sets by comparing the minimum and maximum values of corresponding elements, making it particularly suitable for our data, which includes varying degrees of skill presence. The Weighted Jaccard Similarity metric provides a more accurate measure of similarity between occupations by considering both the presence and relevance of skills, as well as the proportion of common elements relative to the total number in each occupation. By accounting for the number of common skills compared to the total in each occupation, the Weighted Jaccard Similarity offers a more nuanced and realistic representation of occupational similarity.

Using this new metric, we created a similarity matrix that was then transformed using Multidimensional Scaling (MDS), a technique that reduces the dimensionality of high-dimensional data, making it easier to identify patterns and relationships. The MDS transformation provided coordinates in a two-dimensional space for each skill, representing the dimensions of Physical vs. Interpersonal skills. This transformation enabled us to plot the skills and occupations in a manner that visually highlighted their similarities and differences.

For clustering the skills and occupations into meaningful groups, we utilised the k-means clustering algorithm. As shown in Appendix B, Figure B1 and Figure B2, to determine the optimal number of clusters, we used the Elbow Method and Silhouette Analysis. After analysing their plots and conducting several analyses with different valid numbers of clusters, we determined that three clusters provided a good balance between simplicity and explanatory power. Each cluster was then analysed to identify the common characteristics of the skills and occupations they contained.

In a parallel process, we applied a similar methodology to analyse occupations. The occupations were mapped to their required skills and a pairwise similarity matrix was computed using the Weighted Jaccard Similarity metric. The MDS algorithm was again employed to transform these similarities into a two-dimensional space, providing a visual representation of the occupations based on their skill requirements. This visualisation used the same dimensions as the skills analysis, ensuring consistency in interpreting the data.

To enhance the interpretability of the visualisations, we incorporated opacity based on the automation risk scores. Occupations with higher automation risk were displayed with greater opacity, while those with lower risk were more transparent. This added layer of information helped users quickly identify occupations that were not only similar in skills but also less likely to be automated.

This approach provides valuable insights for planning strategies for the training and professional development of users, thus supporting the recommendations mentioned above. In Section 4.3, the results of this procedure will be presented in more detail.

3.4. EVALUATION

In the model performance evaluation phase, we were inspired by the approach outlined by Xu et al. (2023), using precision, recall, and F1 score as evaluation metrics.

Precision is defined as the ratio of true positive predictions to the total number of positive predictions (equation 3.2). It measures the accuracy of the model's predictions for each class. For predicting the importance and frequency of tasks within an occupation, high precision means that the model accurately identifies tasks for each category without incorrectly labelling tasks. This is crucial for ensuring that the model's recommendations are reliable and actionable.

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{3.2}$$

Recall is calculated as the ratio of true positive predictions to the total number of actual positives (equation 3.3). It assesses the model's ability to correctly identify all relevant instances in the dataset. High recall in our case ensures that the model captures all tasks that are indeed important or frequent, which is critical for comprehensive task analysis and resource allocation.

$$Recall = \frac{True\ Positives}{True\ Positives\ +\ False\ Negatives} \tag{3.3}$$

The F1 score represents the combination of precision with recall into a single metric (equation 3.4). It is particularly useful when it is important to balance precision and recall, especially when both false positives and false negatives are equally critical, as is the case in our situation. This metric ensures that the model maintains a balance between precision and recall, providing a more comprehensive view of its performance, while guaranteeing that the model is both accurate and reliable in real-world applications.

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{3.4}$$

In addition, we have included accuracy metric to evaluate our multi-class classification model, as our main objective is to create new variables in ESCO based on existing variables in O*NET. This allows us to compare the performance of different model architectures, ensuring that our improvements to the ESCO database are precise and valid.

Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of instances (equation 3.5). In our context, accuracy provides a quick overview of how well the model performs across all predictions. While BERT's advanced architecture helps in managing some issues of data imbalance, the level of imbalance in our dataset is not severe. This, combined with the application of techniques like data augmentation and class weighting, ensures that accuracy remains a reliable metric for assessing overall performance.

$$Accuracy = \frac{True\ Positives\ +\ True\ Negatives}{True\ Positives\ +\ True\ Negatives\ +\ False\ Positives\ +\ False\ Negatives} \tag{3.5}$$

By integrating these four metrics, we can comprehensively evaluate the model's performance. This ensures that the model not only makes accurate predictions but also makes precise positive predictions, captures all relevant instances, and maintains a balance between precision and recall.

3.5. DEPLOYMENT

In the implementation phase, we aimed to turn all this research into something valuable and tangible for potential users, specifically Generation X professionals in Portugal. This involved developing a web application that integrates a recommendation system that includes all our findings, offering a highly personalised experience adapted to each professional characteristics and personal preferences. This application suggests alternative career paths and provides a detailed guide to the skills needed to facilitate these transitions, supporting users whose occupations are at risk due to advances in artificial intelligence and who may need to transition careers.

Moreover, we leveraged our predictive models to propose improvements to the ESCO database, aiming to achieve the same level of detail as found in the O*NET database.

Finally, we produced this project report to document the entire process in detail, serving as both an academic and scientific contribution.

4. RESULTS AND DISCUSSION

Following the methodology described in the previous section, this segment presents the results. These results are broken down into several parts: the performance of the fine-tuned BERT model, the automation risk scores of various occupations calculated at the task level, the similarity between occupations based on skills, the skill guide for occupation transitions, the analysis of expected and optimal occupation transitions based on skill similarity, and finally, the presentation of the web application that allows the target audience to access all this information.

Thus, we can say that each of these parts covers different aspects of the research questions: What are the potential impacts of AI and automation on occupations for Generation X professionals in Portugal (Section 4.2); what are the expected and optimal transitions (Section 4.5); In what ways can a recommendation system assist Generation X professionals in Portugal in proactively managing and planning their careers in response to AI and automation trends (Section 4.3, Section 4.4, and Section 4.6). These questions guide our exploration of how AI influences job loss and how an intelligent and detailed transition guide can facilitate successful occupation transitions.

4.1. IMPORTANCE AND FREQUENCY OF A TASK

As previously discussed, calculating the risk of automation for an occupation at the task level requires considering three factors: the importance and frequency of the different tasks within an occupation and the level of automation risk that those tasks currently present. In this section, we will present and analyse both the process of obtaining the results and the actual results of the importance and frequency indicators obtained by fine-tuning the pre-trained BERT model.

4.1.1. Model Configuration

In the development phase of the predictive model, the importance and frequency of tasks in various occupations were determined through multiple experiments to optimise the neural network architecture. The model's performance depends not only on the quality of the data but also on the best possible combination of hyperparameters. With the data quality already ensured, as discussed in Section 3.2, our focus shifted to refining the model architecture.

Initially, the performance of our BERT-based-uncased model was suboptimal. With 512 hidden units initially defined, it became apparent that adjustments were necessary due to the excessive complexity of the model, which led to overfitting. We tested reductions to 256,128,64,32 and even 16 hidden units, alongside significant increases in data shuffling, enhancements to the dropout layer, the addition of L2 regularisation, and the creation of special tokens during preprocessing. These changes improved model performance and reduced overfitting, but the problem persisted.

We also applied techniques aimed specifically at minimising overfitting, such as ReduceLROnPlateau to reduce the rate of learning when performance stagnates and cross-validation to using different subsets of data. Despite these strategies aligned with the adjusted architecture, the model's performance did not improve significantly. We tested various batch sizes (64, 32, 16, and 8) and different numbers of epochs (10, 15, and 20).

The small and relatively lower-quality dataset rendered our previous adjustments ineffective. We concluded that the limited amount and quality of data were the primary issues. Initially, we used only O*NET data, which proved insufficient, and we reduced all the class instances to match the class with the fewest values to achieve a balanced dataset. However, once we increased the quantity and quality of the data - leveraging the BERT model’s capability to handle slightly imbalanced datasets through weighting, applying data augmentation and integrating ESCO with O*NET data for the training dataset- we observed a remarkable improvement in the model’s performance. This enhancement made our architectural changes significantly impactful.

To address the dataset limitations further, we reevaluated our cross-validation strategy. Initially, we employed a 3-fold cross-validation approach, which partitioned the dataset into three equal parts. However, we found that allocating only 66% of the data for training within each fold was not sufficient for optimal model performance. Therefore, we adjusted our strategy within the remaining 90% of the data (excluding the 10% reserved for testing), allocating 75% for training and 15% for validation, while maintaining the 3-fold cross-validation, which we concluded was sufficient. This adjustment aimed to maximise the amount of training data while still rigorously validating the model's performance through systematic cross-validation.

We finalised our model with an initial layer of 128 units, which helped manage the model's complexity without overwhelming it. To prevent overfitting, we added L2 regularisation with a factor of 0.01, which acts as a safeguard against the model becoming too accustomed to the training data. The output of this dense layer is then normalised using layer normalisation, which helps the model learn faster and more reliably. After normalisation, we applied a ReLU activation function to introduce non-linearity, allowing the model to understand more complex patterns. Finally, we included a dropout layer with a 40% dropout rate, which randomly sets 40% of the input units to zero during training, further helping to prevent overfitting and making the model more robust. We set the number of epochs to 10 and employed 3-fold cross-validation, as these values proved to be the most effective and sufficient after rigorous testing.

4.1.2. Findings Assessment

4.1.2.1. Training and Validation Dataset

As illustrated in Figure 4.1, the plots show the progression of four key performance metrics across three folds: loss, accuracy, precision, and recall. Each fold consists of training and validation curves, with training represented by the darker colours and validation shown in the lighter colours, providing insight into the model's learning dynamics and generalisation ability.

The first metric, model loss, is shown in the top-left plot of Figure 4.1. We observe a consistent downward trend in the loss values for both training and validation datasets as the number of epochs increases. This indicates effective learning from the training data. Specifically, the training loss decreases steadily across all folds, suggesting that the model fits well with the training data. Similarly, the validation loss also decreases, though it exhibits some fluctuations. These fluctuations reflect the model's response to new, unseen data. The convergence of validation loss towards lower values demonstrates good generalisation performance.

The second metric, model accuracy, is depicted in the top-right plot of Figure 4.1. There is a steady increase in training accuracy across all folds, approaching near-perfect accuracy by the end of the epochs. This trend suggests the model is correctly learning to classify the training data. The validation accuracy also improves over time but levels off and shows some variability. This indicates consistent and reliable performance on unseen data.

The third metric, model precision, is shown in the bottom-left plot of Figure 4.1. Precision increases across all folds, signifying the model's growing ability to correctly identify relevant instances among the predicted positives. Validation precision also improves, although it shows more variability compared to training precision. This variability is expected due to the differences between training and unseen validation data. The overall high precision achieved demonstrates the model's effectiveness in minimising false positives.

The fourth metric, model recall, is illustrated in the bottom-right plot of Figure 4.1. Recall increases across all folds, showing the model's improved ability to identify all relevant instances in the training data. Validation recall also increases with some fluctuations. The upward trend indicates the model's capability to generalise well and capture relevant instances in the validation data, thereby minimising false negatives.

The plots collectively indicate that our model performs well on both training and validation datasets. The steady decline in loss and the rise in accuracy, precision, and recall across all folds signify effective learning and generalisation. While some variability is observed in the validation metrics, this is typical in real-world scenarios where validation data presents new challenges to the model. The close alignment between training and validation metrics suggests the model generalises well without overfitting.

However, a noticeable gap between the training and validation curves persists. This gap can be attributed to the specificity of the task descriptions used in this study. Tasks may appear different due to the varying contexts in which they are described, even if they are practically very similar. This specificity in task descriptions can lead to differences in how the model perceives and processes tasks across different fields, thereby contributing to the observed gap. In any case, this gap is not very significant since, in practice, the data is part of a standard structure for occupations, which users can adjust according to their specific situations through the platform's high level of customisation. Therefore, there is no need to improve the model's performance further, as the few incorrect classifications are not considered serious issues. Furthermore, overall performance is good and consistent across the different folds, which emphasises the robustness of our model, demonstrating that its performance is reliable across different subsets of data and ensuring that our conclusions do not depend on a single data split. Consequently, the quality and reliability of our improvements to the ESCO database are also not compromised by this small gap.

In terms of specific results, we observed that during the initial epoch, the model achieved an accuracy of 51.81% on the training set with a corresponding precision of 57.07% and recall of 40.39%. The validation set showed improved metrics, with an accuracy of 68.49%, precision of 72.82%, and recall of 60.65%. As the training progressed to the final epoch, the model reached a training accuracy of 98.40%, precision of 98.43%, and recall of 98.40%. The validation results for the final epoch showed an accuracy of 89.41%, precision of 89.41%, and recall of 89.41%.

In conclusion, the performance metrics presented in Figure 4.1 demonstrate the efficacy of our model in predicting task importance within occupations. Although Figure 4.1 specifically shows the performance of the model on the importance feature, it is noteworthy that the performance on the frequency feature is very identical, as illustrated in Appendix C, Figure C1. The model achieves high accuracy, precision, and recall while maintaining low loss for both features. These results underscore the model's potential for practical application in upgrading the ESCO database and supporting Generation X professionals in navigating the evolving job market.

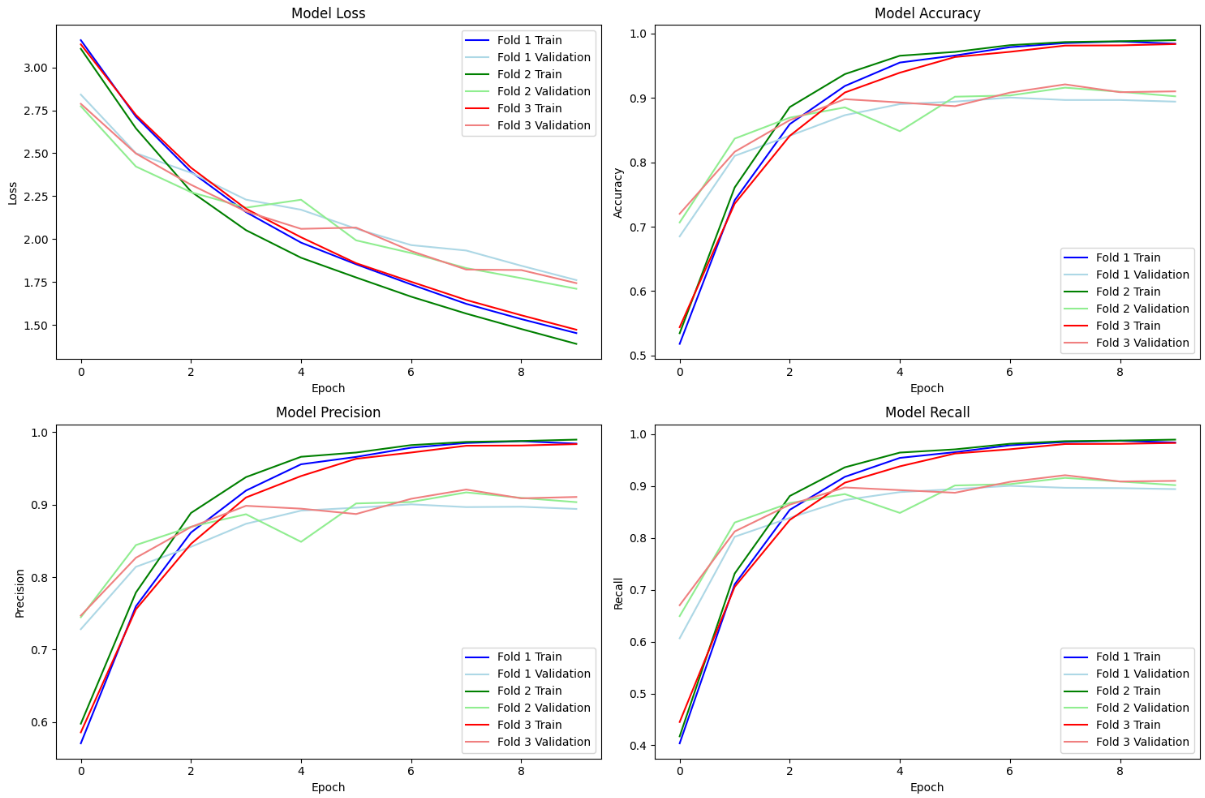

Figure 4.1 – Model Performance Metrics for Importance Prediction

4.1.2.2. Test Dataset

The model was later evaluated using an independent test dataset, which constitutes 10% of the augmented dataset used for training. This portion of the data was kept unseen by the model during training to ensure that our assessment accurately reflects the model's generalisation ability and its potential performance on deployment data. The overall performance metrics are as follows: loss of 1.7395, accuracy of 90.03%, precision of 90.18%, and recall of 89.95%, resulting in an F1 score of 90.06%. These results highlight the model's robustness and its ability to generalise well to unseen data.

The confusion matrix for the importance feature, shown in Figure 4.2, provides a detailed breakdown of the model's predictions versus the actual labels. The matrix reveals the model's ability to correctly classify most instances across all classes. Specifically, for class 0, the model correctly predicted 421 instances, misclassifying 25 instances as class 1 and 4 instances as class 2. For class 1, 273 instances were correctly predicted, with 27 instances misclassified as class 0 and 56 instances as class 2. For class 2, the model correctly classified 354 instances, with only 1 instance misclassified as class 0 and 3 instances as class 1. These results indicate a strong performance in distinguishing between different levels of task importance, with relatively few misclassifications.

From the confusion matrix, we can observe that the model performs better in classifying the extreme classes (0 and 2) compared to the middle class (1). The extreme classes tend to have clearer boundaries, making it easier for the model to classify them correctly. Most of the wrong predictions for these classes are adjacent to them (i.e., class 0 being misclassified as class 1, and class 2 being misclassified as class 1). In contrast, the middle class has more wrong predictions, as it is more likely to be confused with both neighbouring classes. This is evident in the higher number of misclassifications for class 1, which is due to its position in the middle, leading to ambiguity between class 0 and class 2.

In addition to the confusion matrix for the importance feature, we also evaluated the model's performance on the frequency feature. The results for the frequency feature, including the confusion matrix, are presented in Appendix C, Figure C2. The performance metrics for the frequency feature are similarly impressive, with results almost identical to those of the importance feature. This consistency across different features underscores the reliability and versatility of our model.

Overall, the test results demonstrate the model's effectiveness in predicting task importance and frequency within occupations. The high accuracy, precision, recall, and F1 score indicate that the model performs well in real-world scenarios, accurately capturing the nuances of task descriptions and their associated importance and frequency levels.

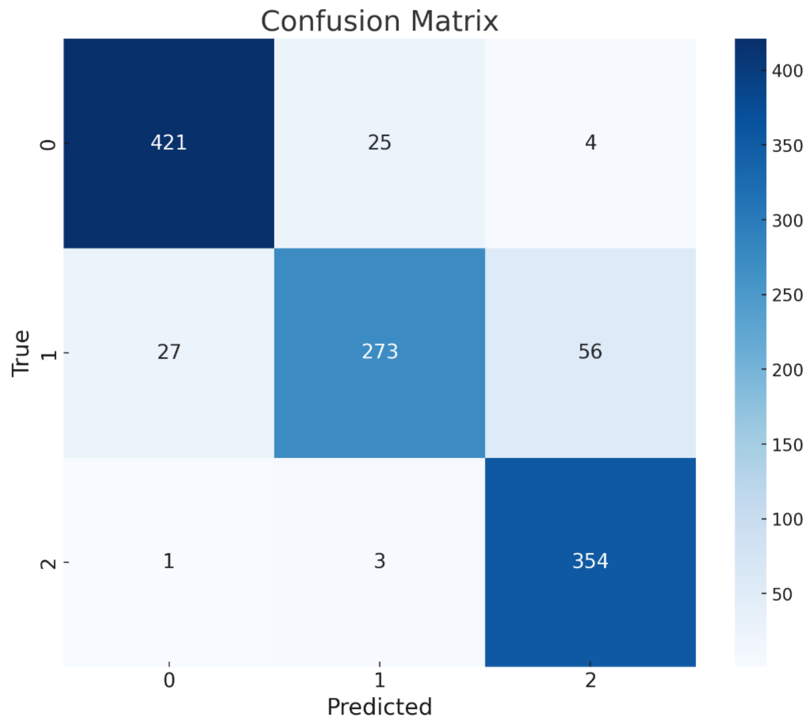

Figure 4.2 – Confusion Matrix of Importance Feature

4.1.3. Deployment Dataset

Having achieved a positive model performance, we proceeded to apply the model to our ESCO dataset. The results indicated that, out of a total of 1257 tasks, 184 tasks were assigned an importance value of 1 (slightly important), 844 tasks were assigned an importance value of 2 (important), and 229 tasks were assigned an importance value of 3 (very important), as illustrated in Figure 4.3.

From the distribution, it is evident that most tasks fall into the moderately important category (class 2), comprising 67.1% of the tasks. The "very important" category (class 3) and the "slightly important" category (class 1) comprise 18.2% and 14.6% of the tasks, respectively. This distribution suggests that most tasks are perceived as having moderate importance, with fewer tasks being categorised at the extremes.

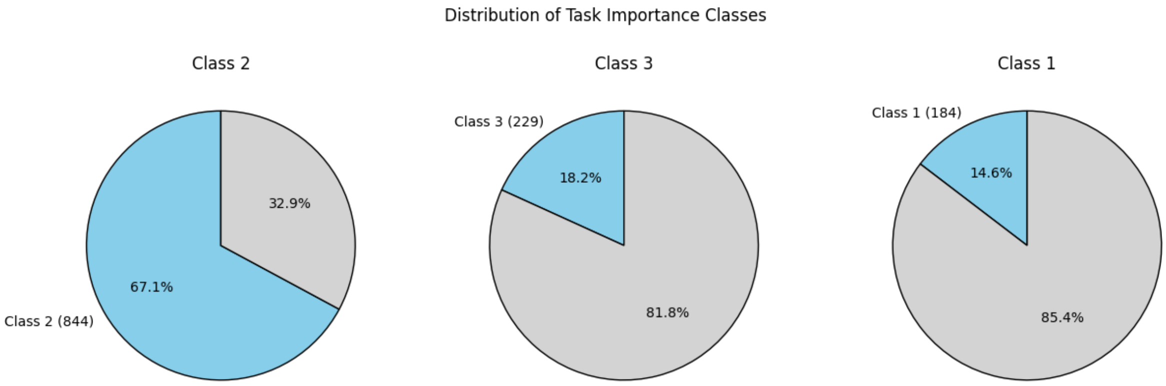

Figure 4.3 – Distribution of Task Importance Classes

We applied the same methodology to the frequency feature, with the results presented in Appendix D, Figure D1. Here, out of a total of 1257 tasks, 250 tasks were assigned a frequency value of 1 (yearly or more), 481 tasks a frequency value of 2 (more than monthly to weekly), and 526 tasks a frequency value of 3 (daily or more frequently). The distribution shows that 41.8% of the tasks are frequent, 38.3% are moderately frequent, and 19.9% are infrequent. This indicates a more balanced distribution across the frequency categories compared to the importance categories.

Regarding the model's confidence in these designations, most tasks were labelled with high confidence. This confidence is crucial for practical deployment, as it suggests reliability in the model's predictions. However, it is essential to remain cautious of potential overfitting, as high confidence in predictions does not always equate to real-world accuracy. Ensuring that the model performs well on unseen data and across different contexts is vital for its successful application.

It is also important to note that even though the model's performance metrics did not vary significantly with different data augmentation techniques, the results on the deployment dataset did show differences in class distribution. This suggests that while augmentation techniques might not drastically impact overall performance metrics, they can influence how the model perceives and classifies tasks in practical applications. This variation in class distribution underscores the need for careful consideration of data augmentation methods in the training process to achieve desired outcomes in deployment scenarios.

In conclusion, the deployment results demonstrate that the model can effectively categorise tasks by importance and frequency. The distributions align with expected patterns, with most tasks falling into moderate categories, which demonstrates the model's greater difficulty in correctly classifying the middle class, as seen previously, which is nevertheless not an exaggeration. These findings highlight the model's potential for practical use in upgrading the ESCO database, providing valuable insights into task importance and frequency that can support Generation X professionals in adapting to changes in the job market. The consistency in model performance across both features further underscores its robustness and reliability for future applications.

4.2. Automation Impact

Once the importance and frequency of the tasks have been obtained and explained, we proceed to describe the results in terms of the risk of automation that these tasks present. As mentioned previously, this risk was obtained through requests to GPT-4. Below, we present the results of this assessment, as well as the resulting insights at the occupation level.

4.2.1. Task Level

The analysis of the risk distribution of automation scores at the task level within occupations, as illustrated in Figure 4.4, highlights the differences between the averages of three distinct occupational categories: Professionals, Service and Sales Workers, and Elementary Occupations. The coloured vertical lines represent the averages of each major group, showing that elementary occupations have a higher susceptibility to automation, followed by service and sales workers, and, finally, professionals, which exhibit the lowest average of automation. This distribution provides a comprehensive view of the potential for automation in different sectors of the labour market, suggesting which categories may face significant changes soon due to the advancement of technology.

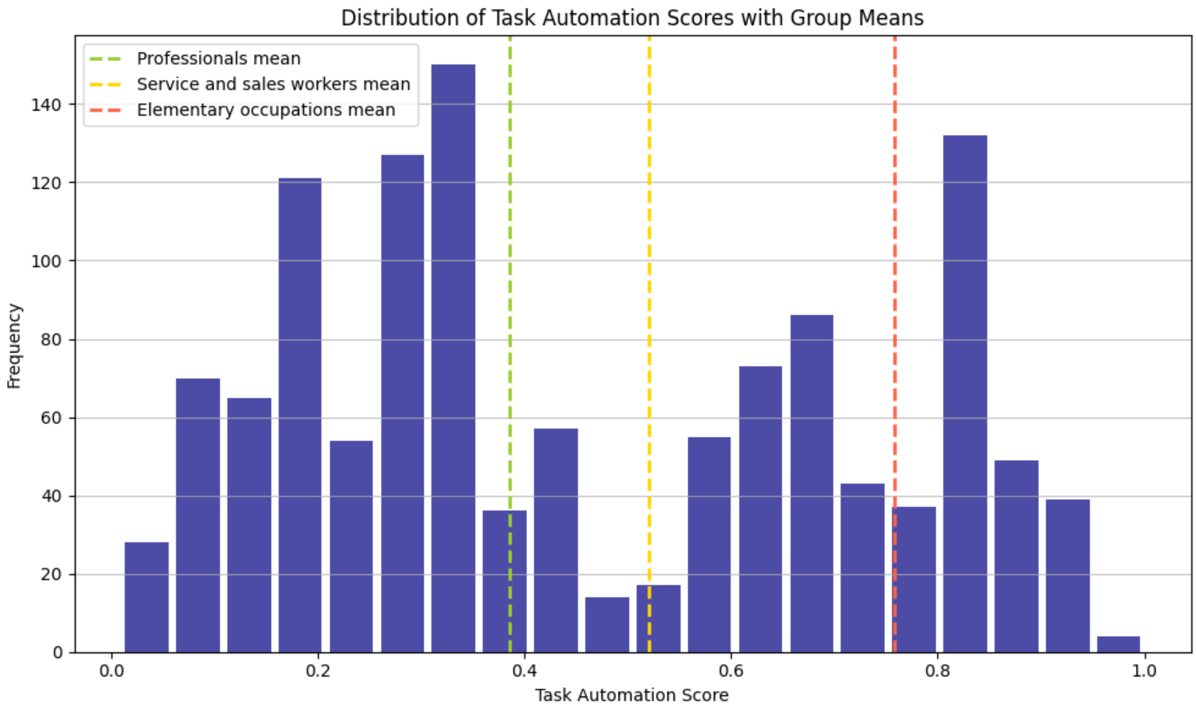

Figure 4.4 – Distribution of Task Automation Scores with Group Means

Figure 4.5 indicates that tasks with the highest risk of automation have probabilities nearing 100%, characterised by routine and predictable activities, which justifies their high percentage of automation risk. On the other hand, tasks with lower risk have values close to 0%, indicating a greater need for human skills to perform them. Among the high-risk tasks, activities such as “opening and closing doors for passengers” and “washing and polishing vehicle windows” stand out. In contrast, tasks such as “propagating religious doctrines in own country or abroad” and “training, exercising and attending dance classes to maintain the required levels of ability and fitness” are among those with the lowest risk of automation.

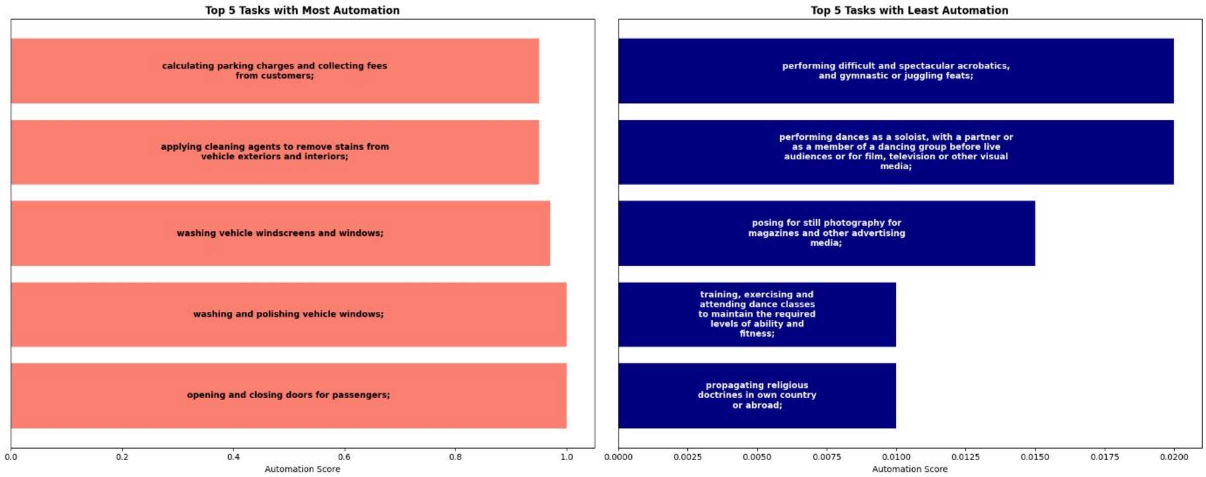

Figure 4.5 – Top 5 Tasks with Most and Least Automation Risk

Comparing our results with the findings of Lassébie and Quintini (2022), who used skills and abilities to measure the degree of automatability of an occupation, reveals a consistent trend in automation risk. Both studies indicate that activities involving complex human interactions, creativity, and personal judgment are less susceptible to automation. In the Lassébie and Quintini study, skills such as “negotiation”, “social perceptiveness”, and “active listening” are identified as having low automation risk. Similarly, in our study, tasks like “performing difficult and spectacular acrobatics, and gymnastic or juggling feats” and “posing for still photography for magazines and other advertising media” demonstrate low automation risk.

On the other hand, both studies identify repetitive, routine, and physically intensive activities as having higher automation risk. Lassébie and Quintini's findings highlight skills such as “speed of limb movement” and “digital data processing” as high risk, which aligns with our identification of tasks like “washing vehicle windscreens and windows” and “calculating parking charges and collecting fees from customers”. This alignment underscores a broader trend where the nature of work—whether it requires human touch and creativity or involves repetitive actions—significantly influences its automation risk.

4.2.1.1. Occupation Level

After examining the analysis of the tasks most and least susceptible to automation, we can further analyse the dimension of occupations. Figure 4.6 presents a broader analysis, showing how the level of automation risk is distributed across three major occupational groups:

Elementary Occupations, Professionals, and Service and Sales Workers. The figure includes four defined exposure thresholds, marked with horizontal lines: Very low risk for scores below 0.25, low risk for scores from 0.25 to just under 0.50, medium risk for scores from 0.50 to just under 0.75, and high risk for scores of 0.75 and above.

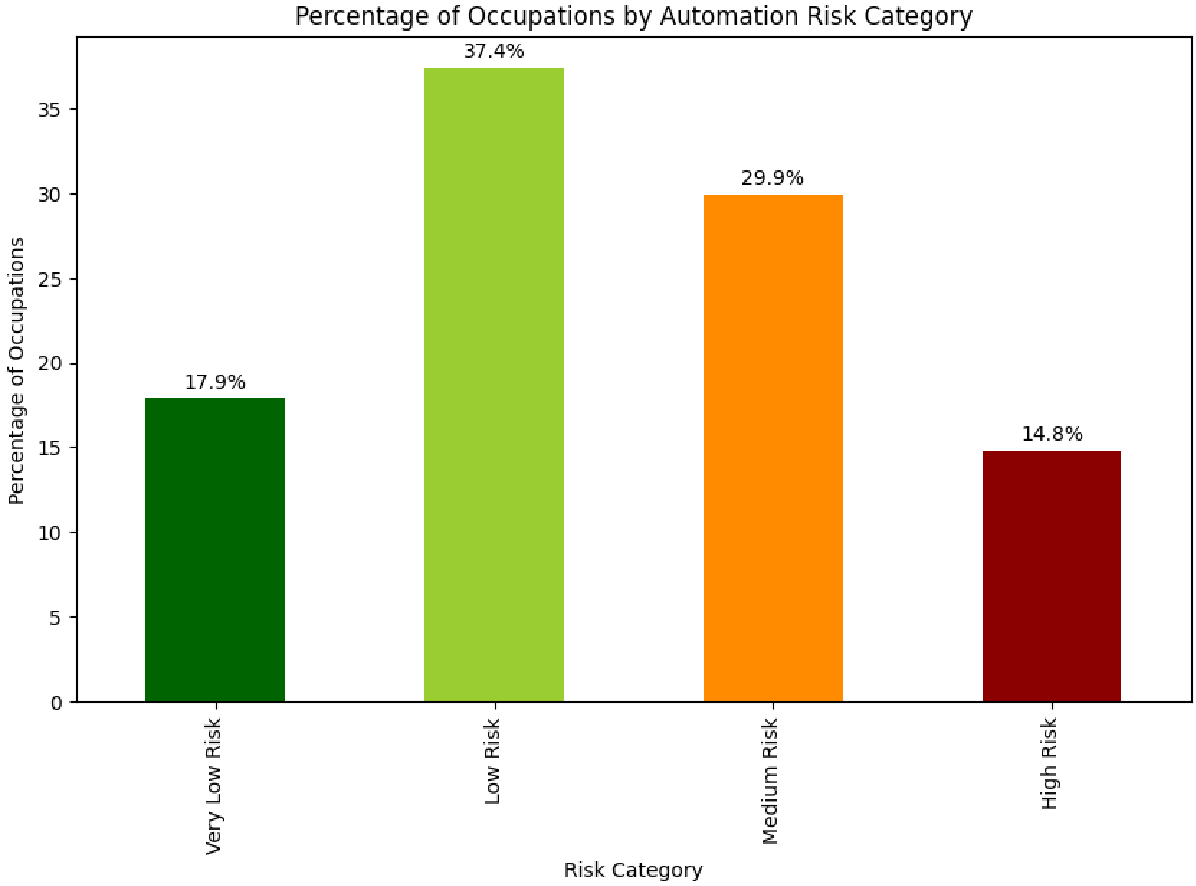

Figure 4.6 – Percentage of Occupations in Automation Risk Category

Figure 4.6 demonstrates the distribution of every 4-digit occupation analysed within those thresholds. Most occupations, accounting for 37.4%, fall into the Low-Risk category, indicating that these jobs have a moderate likelihood of being automated. Medium-Risk occupations make up 29.9%, suggesting that a significant portion of jobs have a reasonable chance of automation. Very Low-Risk jobs comprise 17.9%, showing that a smaller fraction of occupations is relatively safe from automation threats. Lastly, High-Risk occupations account for 14.8% of the total, highlighting that a notable percentage of jobs are at immediate and high risk of automation.

It is difficult to compare this overall perspective of our findings with existing literature because other studies analyse all occupational groups, whereas we, mainly due to computational power and costs, focus only on the three most frequent occupational groups of Generation X in Portugal. Consequently, we cannot directly analyse groups with high exposure to automation risk that are included in other studies, such as Technicians and Associate Professionals and Clerical Support Workers (Gmyrek et al., 2023). Our study serves more as an introduction to a new methodology and an innovative approach to providing personalised access to this data through the development of a recommendation system integrated into a web application, designed for stakeholders while being scalable in the future not only in terms of occupations analysed but also across several countries.

As seen before, elementary occupations have the highest average risk of automation at 76%, followed by service and sales workers with an average risk of 52%, and professionals with an average risk of 39%. This distribution highlights the different levels of susceptibility to automation across different occupational categories. As Figure 4.7 illustrates, for elementary occupations, most of the 4-digit occupations fall into the medium and high-risk categories, with 26.9% of occupations classified as medium risk and 65.6% as high risk. Only 7.5% of occupations fall into the low-risk category, and there are no occupations in the very low-risk category. This makes elementary occupations more susceptible to automation compared to the other two groups.

In contrast, the professional category shows a significant portion of occupations in the low to medium-risk categories. Specifically, 23.8% of occupations are classified as very low risk, 50.5% as low risk, and 23.8% as medium risk. A small percentage, 1.9%, falls into the high-risk category.

Service and sales workers show a varied distribution of automation risk. Specifically, 15.1% are classified as very low risk, 23.5% as low risk, 49.6% as medium risk, and 11.8% as high risk. This distribution indicates that while some occupations in this category are highly automatable, the overall risk remains moderate. The risk distribution for each group reflects the complexity and variability of tasks within those occupations, affecting their susceptibility to automation.

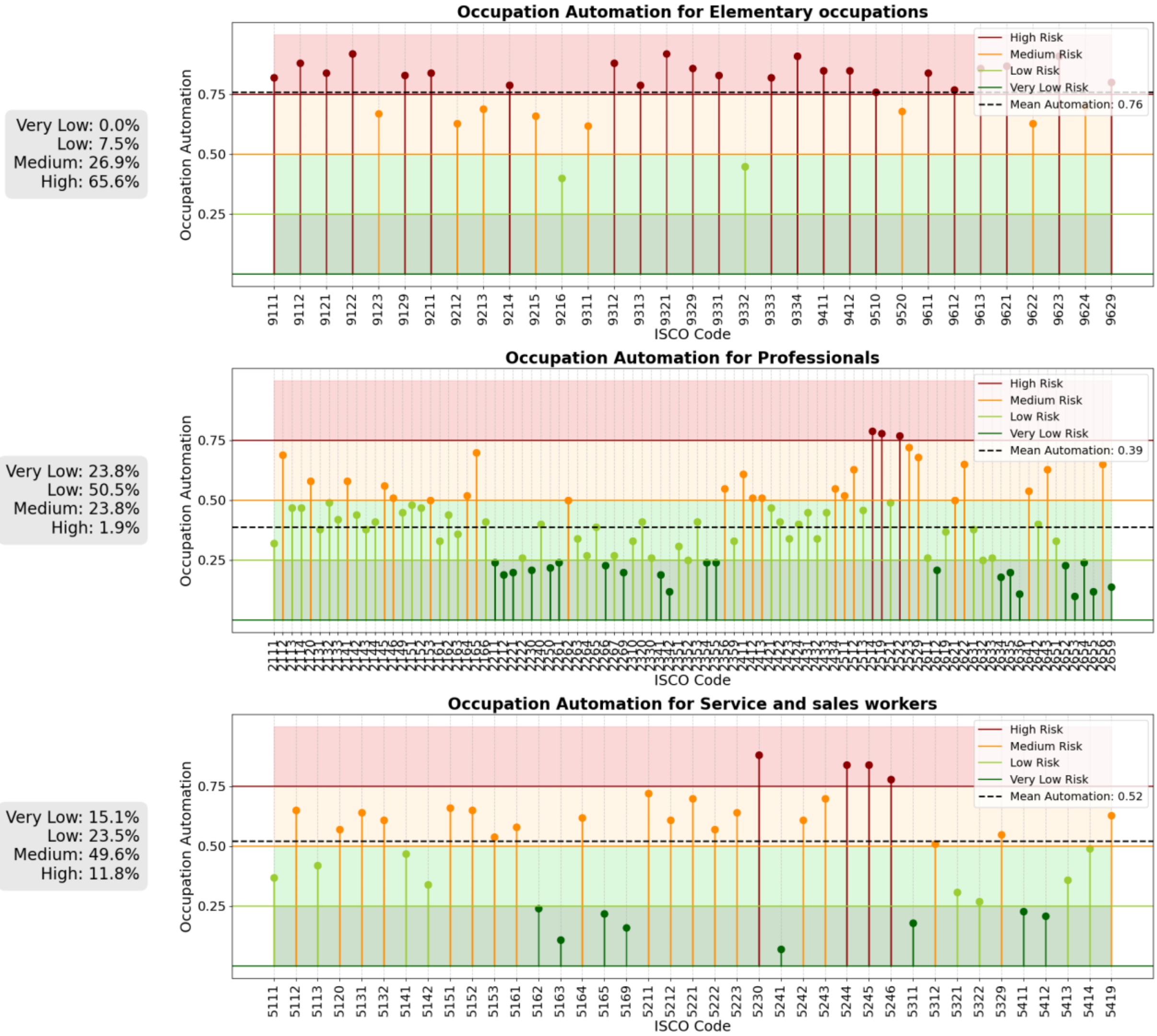

Figure 4.7 – Risk of Automation along Major Occupation Group

Building on the insights from Figure 4.7, which highlighted the varying levels of susceptibility to automation among the three major occupational groups (1-digit ISCO code), Figure 4.8 further illustrates the distribution of exposure tasks among these groups. This figure shows that tasks with a high risk of automation represent around 30% for service and sales workers, 8% for professionals, and 72% for elementary occupations. When examining tasks with a medium risk of automation, the data reveals that 24% of tasks for service and sales workers have medium exposure, compared to 20% for professionals and 16% for elementary occupations. Notably, professionals have an additional 28% of tasks categorised as very low risk.

These findings support the conclusions from Figure 4.4, reinforcing that elementary occupations face the highest average risk of automation among the three occupational groups. The data underscores the varying degrees of automation risk each group faces, highlighting the greater vulnerability of elementary occupations to automation compared to service and sales workers and professionals.

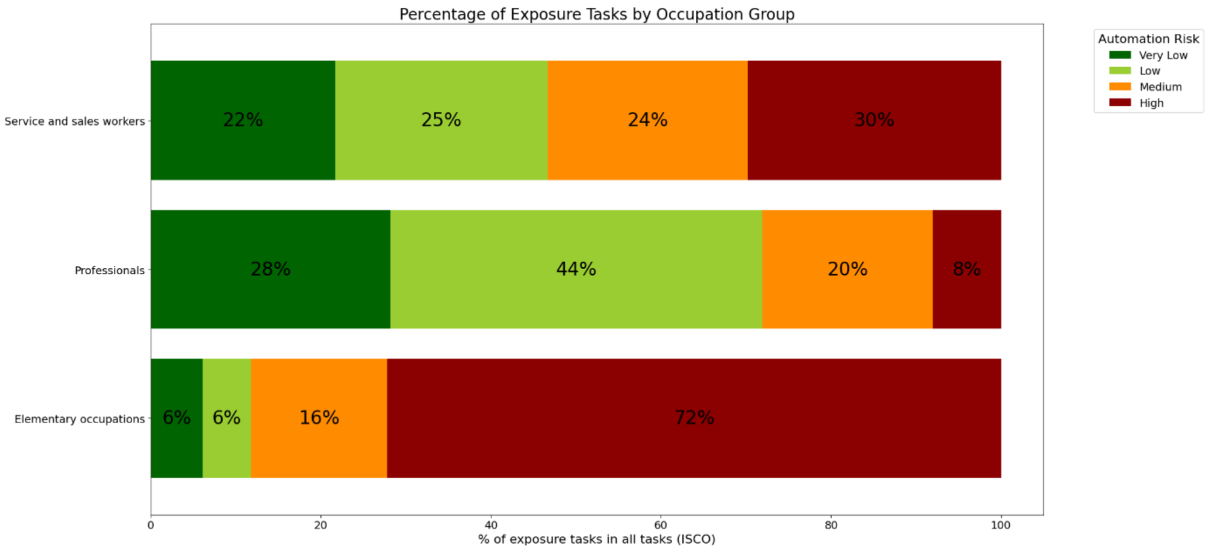

Figure 4.8 – Percentage of Exposure Tasks by Occupation Group (1-digit)

By comparing the results with those of Gmyrek et al. (2023) for these three specific groups, we realise that there is a general increase in the risk of automation across occupation groups. The risk of automation has increased across the distribution of occupations in the different groups. For elementary occupations, Gmyrek et al. (2023) obtained an average level of exposure to automation of less than 25%. Our results, on the other hand, show an increase to 76% (Figure 4.7), with most occupations exceeding the 75% automation risk threshold, entering the high level. This did not occur in any occupation in this group according to Gmyrek et al. (2023).

In the professionals and service and sales workers groups, Gmyrek et al. (2023) already identified a large proportion of occupations at medium and high risk. Our results show a slight increase across these two groups, which can be attributed to the technological advancements from 2023, the year of Gmyrek's study, to the current year of our project.

Analysing the tasks covered by each of these groups, we observe notable changes in the percentage of tasks at high risk of automation. For professionals, the high-risk percentage increased from 1% to 8%, for service and sales workers, it rose from 4% to 30%, and for elementary occupations, it surged from 1% to 72%. In terms of tasks with a medium level of automation risk, professionals saw a decrease from 25% to 20%, which can be attributed to some tasks migrating to the high-risk category. Service and sales workers experienced an increase in medium-risk tasks from 18% to 24%, and elementary occupations went from 3% to 16%. These changes confirm a general rise in the risk of automation across various 1-digit occupational groups.

The primary reason for this discrepancy, besides the rapid evolution and increasing power of AI, is that Gmyrek et al. (2023) measured the potential for automation only using GPT. In contrast, our study incorporates both GPT and additional AI technologies, leading to a more comprehensive assessment of automation risks.

Similar to Figure 4.5, in Figure 4.9 we observe a clear distinction in automation risk among different occupations. On the one hand, occupations such as “Civil engineering labourers”, “Shelf fillers”, “Meter readers and vending-machine collectors”, “Vehicle cleaners”, and “Hand packers” are the most susceptible to automation. On the other hand, occupations such as “Actors”, “Undertakers and embalmers”, “Religious professionals”, “Dancers and choreographers”, and “Fashion and other models” are among the least vulnerable to automation. The low automation risk score for these occupations indicates a lower probability of the tasks being automated, largely due to their high complexity or the need for human interpretation and interaction.

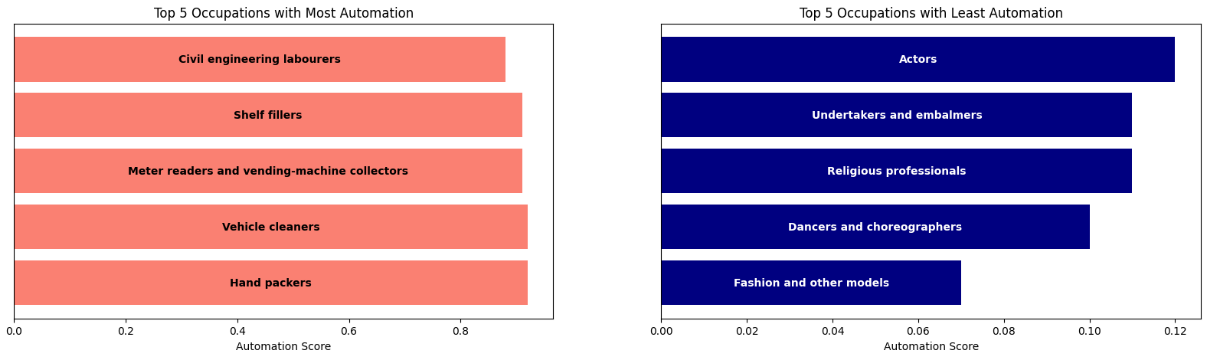

Figure 4.9 – Top 5 Occupations with Most and Least Automation

Lassébie et al. (2022) measure the degree of automatability by summing the product of automatability and importance across skills and abilities. In contrast, we used a more complex formula that incorporates frequency and evaluates tasks instead of skills, providing a more detailed risk assessment. Despite these methodological differences, both Lassébie and Quintini (2022) and our findings reveal similar perspectives on which occupations are most and least susceptible to automation. Physical and manual labour-intensive roles, such as construction, extraction, and transportation, have the highest risk. Conversely, roles requiring complex decision-making and human interaction, like management, education, and legal occupations, are deemed the least automatable.

Building on this discussion, Xu et al. (2023) provide an analysis at a more detailed level of individual occupations. While our analysis focuses on the 4-digit occupational group level, both studies identify functions involving repetitive and routine tasks as the most automatable.

4.3. Similarities Across Occupations - Cluster Analysis

After exploring the impact of automation from individual tasks to entire occupations, it is crucial to understand the relationships between skills within different occupations and the occupations themselves. By examining these relationships and the relevance of each skill, we can facilitate potential transitions between occupations. This approach will not only help assess the importance of skills within occupations but also provide insights into the transferable and adaptable skills needed for smooth and successful transitions, effectively supporting individuals in navigating career changes.

4.3.1. Skill and Competence Level

In this section, we analyse the similarities and differences across different skill levels, utilising a framework of dimensions and clusters. This analysis is grounded in the definitions of the dimensions represented on the x and y axes. The dimensions provide a spectrum that categorises tasks based on the extent of physical versus cognitive skills and individual versus interpersonal skills.

The dimension represented on the x-axis is "Physical" which captures the essence of practical, hands-on, and physical skills on one end, and abstract, conceptual, and cognitive skills on the other. The y-axis dimension is "Interpersonal" representing skills that involve collaboration, teamwork, and social interaction versus those that are more individual-focused and less collaborative.

Three clusters emerged from this analysis, each representing a unique combination of these dimensions. The "Interpersonal Skills" cluster is characterised by skills that are highly collaborative and involve social interaction. This cluster, predominantly located in the upper half of Figure 4.10 and coloured in green, includes skills such as "solving problems", "applying civic skills and competences", and "developing financial, business, or marketing plans". These skills emphasise teamwork and interaction with others.

The "Physical Skills" cluster, illustrated in orange, involves practical, hands-on skills that are more individualistic and less focused on collaboration. Found primarily in the right half of the plot, with values concentrated lower on the y-axis (interpersonal dimension), these skills require physical effort and practical abilities. Examples include "applying material to fill gaps in surfaces", "building and repairing structures", and "installing interior or exterior infrastructure". These skills are oriented towards direct and physical activities.

The purple cluster represents "Cognitive Skills", which encompasses skills that are more abstract, conceptual, and cognitive in nature. Concentrated in the lower-left quadrant, with values located further down in the interpersonal dimension, this cluster includes skills like "fitting assistive devices", "advising and consulting", and "providing personal care". These skills are defined by their cognitive demands and less emphasis on physical interaction.

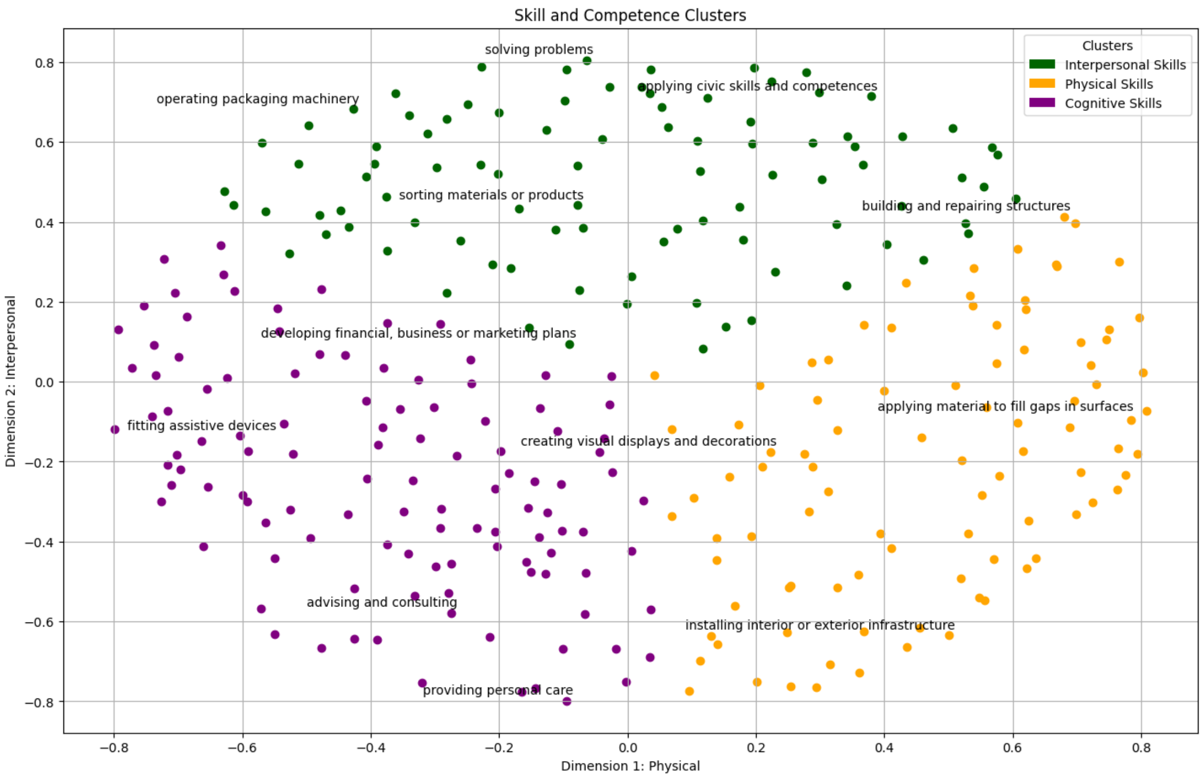

Figure 4.10 – Skill and Competences Clusters Based on Physical and Interpersonal Dimensions

Figure 4.11 further illustrates the distribution of different skill categories across the clusters, providing a detailed view of how various skills are grouped. This histogram highlights the predominant skill categories within each cluster, emphasising the diversity and concentration of skills. For instance, the "Interpersonal skills" cluster shows a significant concentration of skills in "communication, collaboration and creativity", as well as in "working with machines and specialised equipment", which proves its distribution along the x-axis (physical dimension) indicating that the cluster covers both the cognitive and physical environments, being the only cluster to have the presence of very skill category. Additionally, the predominance of the “handling and moving” category within this cluster is evident, reflecting the density of skills in the positive x-axis.

The “Physical Skills” cluster is dominated by “handling and moving”, “working with machines and specialised equipment”, and “constructing” skills, indicating a focus on practical, hands-on tasks. Meanwhile, the “Cognitive Skills” cluster has a notable presence of “communication, collaboration, and creativity”, “information skills”, “assisting and caring”, and “management skills” underscoring the cognitive and creative essence of these skills.

This distribution highlights the unique characteristics and requirements of each skill cluster, providing valuable insights into the nature of these skills.

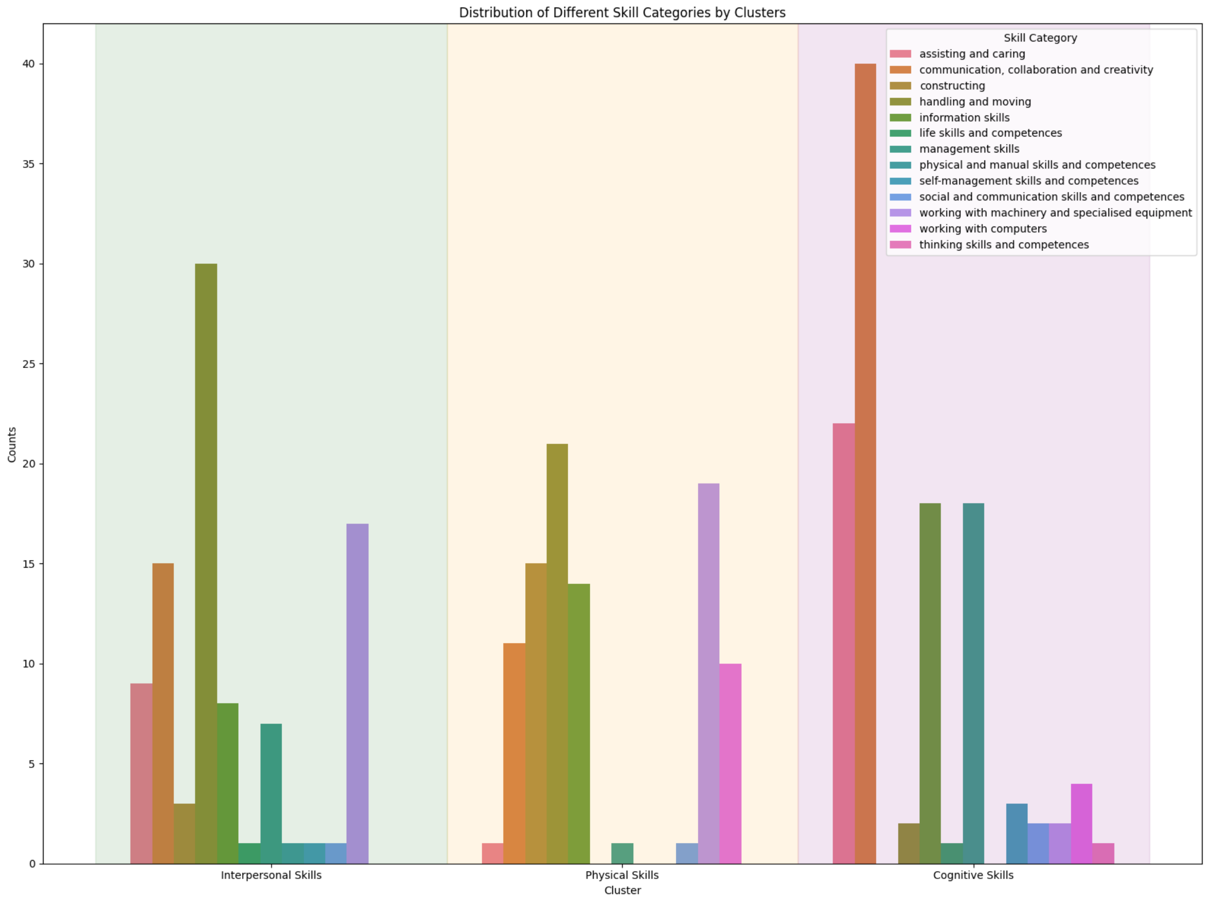

Figure 4.11 – Distribution of Different Skill Categories by Clusters

This analysis not only categorises skills based on their levels and competencies but also provides a visual representation of their distribution along two critical dimensions. The coherence between the dimensions, cluster definitions, and their positioning on the plot underscores the robustness of this framework.

4.3.2. Occupation Level

Building upon our previous examination of skill and competence levels, we now shift our focus to exploring the similarities and patterns observed across various occupational roles. This analysis leverages the same dimensions of "Physical" and "Interpersonal" to categorise occupational roles, thereby providing a coherent framework that links skills to occupational clusters.

In the skill analysis, we identified three distinct clusters, each characterised by a unique combination of these dimensions. Similarly, in our occupational analysis, we observed comparable clusters that align with the identified dimensions. This alignment not only validates the dimensional framework but also elucidates the nature of different occupational roles and the skills and competences they demand.

Cluster "Interpersonal Occupations" includes occupations that heavily rely on interpersonal skills. Examples of such occupations are "Actors" and "Dieticians and Nutritionists" indicating roles that require teamwork and interaction. These roles are primarily located in the upper half of Figure 4.12, where interpersonal skills are predominant, and are represented in green.

Cluster “Cognitive Occupations” encompasses occupations that require significant cognitive skills and are often associated with problem-solving and less physical activity. Occupations in this cluster, such as “Applications Programmers”, “Social Work and Counselling Professionals”, and “Vocational Education Teachers”, combine cognitive efforts with fewer interpersonal interactions. Positioned in the lower-left quadrant, these roles are more about cognitive tasks and less about physical activities. This cluster is shown in purple.

Cluster “Physical Occupations” consists of occupations necessitating a high degree of physical skills. Roles like “Vehicle Cleaners”, “Hand Packers”, and “Domestic Housekeepers” fall into this cluster, emphasising physical activity within structured contexts. These occupations are concentrated in the lower-right quadrant, reflecting the need for precise execution of physical work. This cluster is illustrated in orange.

The opacity of each point in Figure 4.12 represents the occupation's automation risk, highlighting the varying levels of susceptibility to automation across different roles. Occupations with higher automation risk, such as “Applications Programmers” are represented with more opacity and occupations with lower automation risk, such as “Actors” are represented with less opacity.

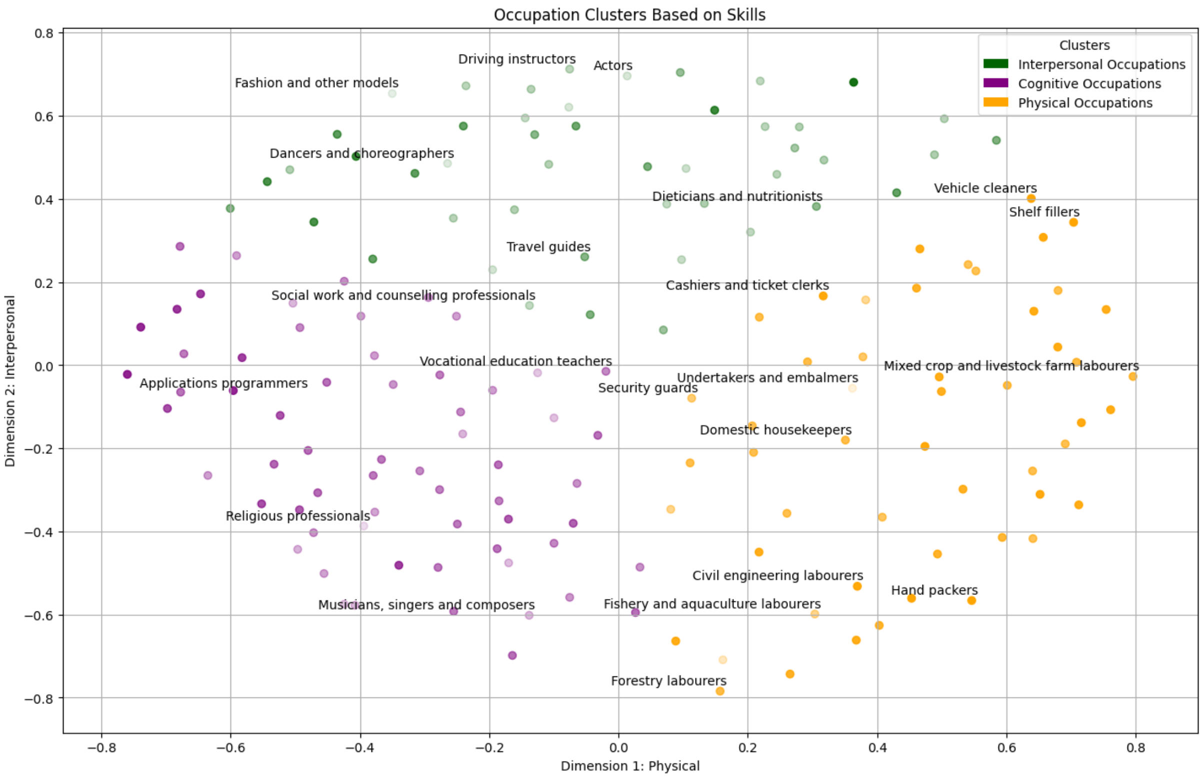

Figure 4.12 – Occupation Clusters Based on Physical and Interpersonal Dimensions

As we can see in Figure 4.13, the distribution of major occupation groups across these clusters further underscores the alignment of certain occupation categories with specific skill dimensions. For instance, the “Cognitive Occupations” cluster includes only occupations from the professional group, indicating that this group is characterised by a higher skill demand and more complex occupations, such as “Application Programmers” and “Vocational Education Teachers.”

The “Interpersonal Occupations” cluster includes many occupations from the service and sales workers group, emphasising the collaborative and interpersonal nature of these roles. Although there is a predominant presence of occupations from the professional’s group in this cluster, this can be attributed to the larger number of occupations analysed from this occupational group compared to the other two major groups (service and sales workers and elementary occupations).

In contrast, the “Physical Occupations” cluster includes a mix of roles from the service and sales workers and elementary occupations groups, with almost no representation from the professionals group. This is significant especially given the disproportionately large number of occupations analysed from the professional group, highlighting the lack of physical nature in occupations from the professionals group. This underscores the technical nature of those roles and the physical nature present in occupations from the service and sales workers and elementary occupations groups, such as “Domestic Housekeepers”, “Travel Guides”, ”Vehicle Cleaners”, and “Garbage and Recycling Collectors”.

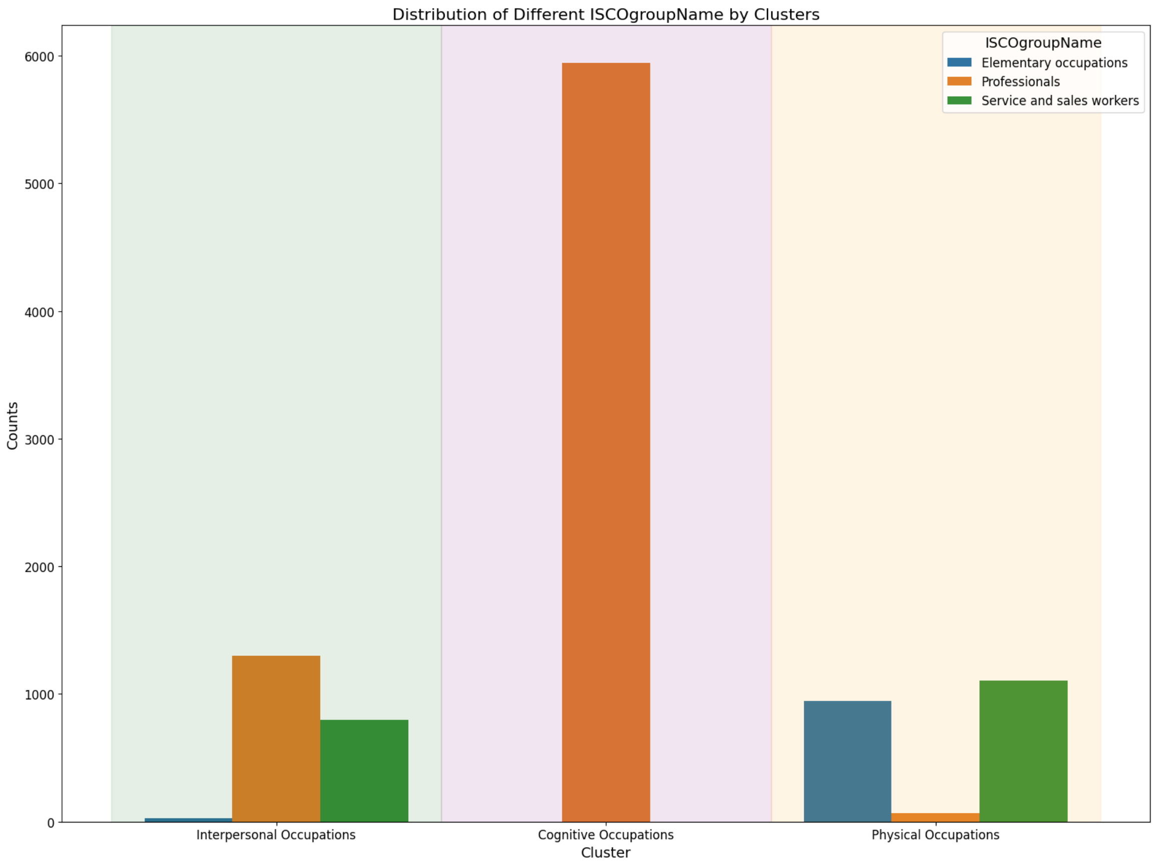

Figure 4.13 – Distribution of Different Isco Groups by Clusters

The coherence between the dimensions and clusters, and the illustrative examples, highlights the exchange between individual initiative and collaboration as well as between technical and physical efforts across different job roles and skill sets. This integrated view of skills and occupations, grounded in empirical data and visualised through cluster analysis, offers a comprehensive understanding of the competency landscape. This understanding is essential for both workforce planning and individual career development.

By comprehending these clusters and their characteristics, we can tailor training programs, occupational roles, and performance evaluations to better align with the nature of the tasks involved, thereby enhancing overall efficiency and effectiveness in various professional contexts.

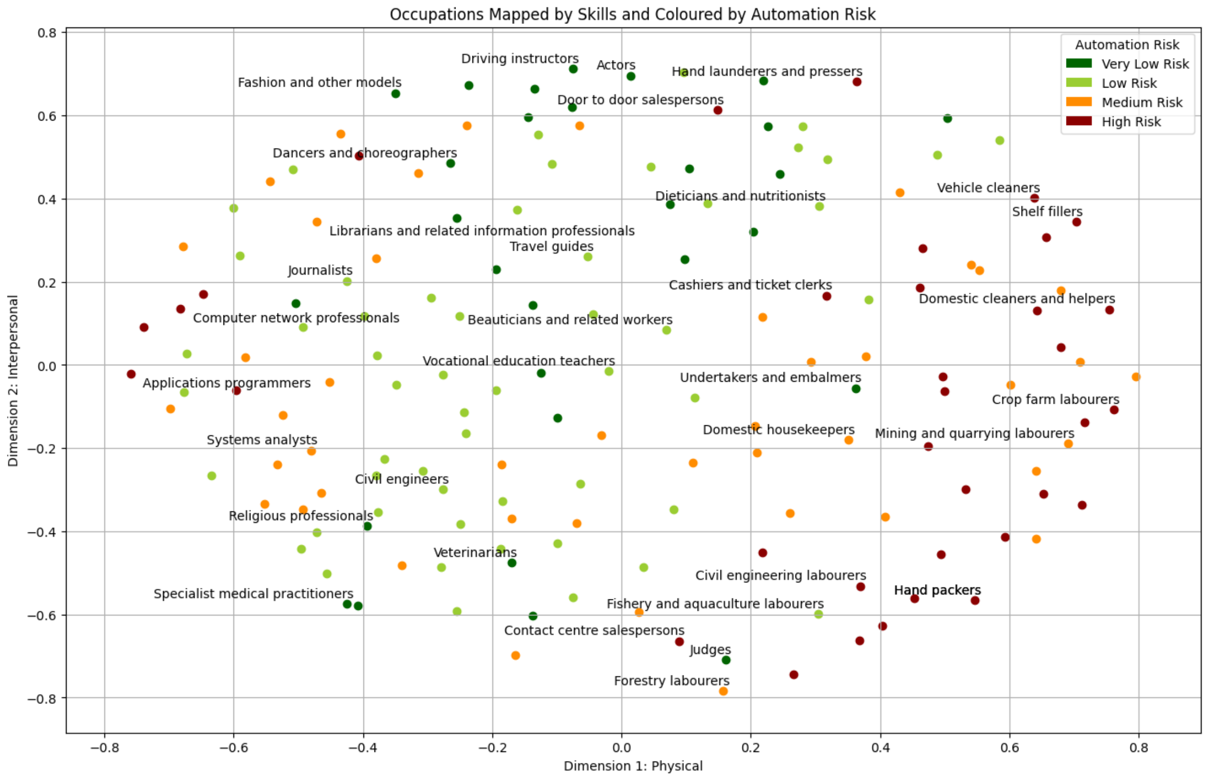

Figure 4.14 – Distribution of Different Occupations by Skills and coloured by Automation Risk

Figure 4.14 illustrates the distribution of occupations across the two dimensions, with colour coding representing automation risk categories from Very Low Risk (dark green) to High Risk (dark red). This visualisation allows us to identify patterns in how different occupations are positioned concerning their susceptibility to automation based on their skill requirements and task nature.

Occupations requiring physical skills, indicated by positions on the right side of the plot (positive values on Dimension 1), exhibit a notable concentration in the medium-risk and high -risk categories. For example, occupations such as “Mining and quarrying labourers”, “Hand packers”, and “Shelf fillers” which involve significant practical tasks and minimal technical effort, are at higher risk of automation. Their automation scores of 0.62, 0.92, and 0.91, respectively, reflect the routine and repetitive nature of their tasks, which can be more easily automated.

On the other hand, occupations demanding higher technical skills, located on the left side (negative values on Dimension 1), such as “Fashion and other models”, “Specialist medical practitioners”, and “Civil engineers”, with automation scores of 0.07, 0.19, and 0.44, respectively, tend to fall into the very low-risk to low-risk categories, indicating less susceptibility to automation in occupations with a more technical environment.

Regarding the Interpersonal dimension (Dimension 2), occupations involving fewer interpersonal skills, found at the bottom (negative values), are often in the medium-risk to high-risk categories. For instance, “Forestry labourers” (automation score 0.66) and “Contact Centre Salespersons” (automation score 0.84) highlight the vulnerability of solitary labour roles to automation. On the other hand, occupations that involve interpersonal skills, positioned towards the top (positive values), such as “Actors” and “Driving Instructors” (with automation scores of 0.12 and 0.22), are generally less prone to automation, falling into the “Very Low Risk” to “Low Risk” categories, demonstrating the major difficulty of AI to replicate human interactions.

Additionally, “Vocational education teachers” and “Beauticians and Related Workers” (with automation scores of 0.41 and 0.34) are centrally positioned, reflecting a balance of cognitive skills and some technical skills, with a lower risk of automation due to the need for human interaction and adaptability.

Our results align with the conclusions of previous research (Brynjolfsson & McAfee, 2011, 2014; Nedelkoska & Quintini, 2018; Xu et al., 2023), indicating a substantial risk of job displacement in sectors dominated by repetitive and low-skilled tasks. These studies, particularly Xu et al. (2023), emphasise the vulnerability to automation of occupations that require basic pattern recognition, such as dishwashers and data entry clerks. Our findings support these conclusions by demonstrating that, over time, not only low-skilled tasks with basic pattern recognition are the most affected, but also more complex and higher-skilled tasks are increasingly at risk. This trend places occupations requiring technical skills, such as “Application programmers” and “Computer network professionals”, at high risk of automation. These occupations involve complex pattern recognition. As a result, tasks previously seen as bottlenecks are becoming automatable (Lassébie and Quintini, 2022; Gmyrek et al., 2023), which indicates that as automation technologies advance, the impact of automation is beginning to spread to more highly qualified occupations that require more technical skills and more complex cognitive tasks (Giattino et al., 2023).

In conclusion, the plot reveals that occupations requiring high physical skills are generally at a higher risk of automation, while those involving technical skills are becoming more susceptible over time. This trend can be attributed to advancements in robotics and machine learning, which initially could more easily replicate physical and routine technical tasks but now are becoming more advanced and starting to replicate complex cognitive functions, although replicating human interactions remains challenging (Giattino et al., 2023). Therefore, developing skills that leverage human communication, collaboration, and creativity capabilities is crucial for reducing automation risks.

4.4. Recommendation System

Having gathered the automation risk scores for each occupation and measured the skill distances between them, we can now build our recommendation system. This system will suggest occupations with lower automation risk and similar skill sets to the user's current occupation. Based on the user's selection, the recommendation system will provide a skill guide to facilitate a smooth occupation transition.

For example, consider the transition from Software Developers to Web and Multimedia Developers. Software Developers face a medium automation risk of 63%, whereas Web and Multimedia Developers have a significantly lower automation risk of 46%, with a similarity score of approximately 51%. This transition is notable as it represents a viable pathway within optimal transitions, focusing on target occupations with less than 50% automation risk.

The skill guide categorises the required skills for the new occupation into three distinct groups.

First, the high-importance skills are crucial in the target occupation and significantly different from those in the source occupation. Users should prioritise developing these skills as they are essential for success in the new role. For this transition, high-importance skills include artistic and creative writing, browsing, searching and filtering digital data, and developing financial, business, or marketing plans. These skills are essential for excelling in the new role and should be the primary focus during the transition.

Second, the low-importance skills have lower importance in both the source and target occupations. While it is beneficial to maintain these skills for a well-rounded skill set, they should not be the primary focus during the transition. Examples include analysing business operations, coordinating activities with others, and documenting and recording information.

Finally, the low-distance skills exhibit minimal differences in importance between the source and target occupations, indicating areas of compatibility. Skills like complying with legal and organisational guidelines, conducting investigations, and developing operational policies and procedures fall into this transition. Users should leverage these familiar skills as a solid foundation for their transition.

By focusing on these categorised skills, illustrated in Appendix E, Figure E5, the recommendation system ensures that users can effectively prepare for their new occupation, minimising the impact of automation risks and enhancing their career stability.

4.5. FUTURE TRANSITIONS

4.5.1. Expected Transitions

In this chapter, we explore the expected transitions between occupations based on skill similarity, focusing on shifts where the target occupation has a lower automation risk (less than 50%), and the base occupation has a medium or high automation risk (50% or more). We consider valid transitions to be those where there is at least a 25% similarity in skills between the base and target occupations. This analysis provides insights into potential career pathways for workers in high-risk occupations, highlighting more secure and sustainable job roles they can transition into.

Figure 4.15 showcases these transitions using the same dimensions and clusters as in previous figures. Transitions are represented by lines connecting occupations, indicating pathways from high-risk to low-risk roles based on skill similarity. A total of 37 transitions were identified, with similarity scores ranging from approximately 25% to 66%, and an average of 40%, indicating a moderate level of skill overlap.

Key observations include transitions such as "Database Designers and Administrators" moving from high-risk occupations like "Applications Programmers", "Computer Network Professionals", and "Systems Administrators" to lower-risk occupations within the same cognitive domain. Another example is "Economists" transitioning from "Authors and Related Writers", "Financial Analysts", "Mathematicians", "Actuaries and Statisticians", and "Town and Traffic Planners", reflecting a shift within analytical and financial fields that leverages their quantitative skills. Additionally, since we are analysing the expected transitions and not the optimal ones, we see tougher transitions as indicated by the length of the vectors in Figure 4.15. These longer vectors signify a greater distance in skill similarity, such as those for "Technical and Medical Sales Professionals (excluding ICT)" transitioning from roles like "Cashiers and Ticket Clerks", "Information and Communications Technology Sales Professionals," and "Shop Keepers", showcasing a pathway that capitalises on their sales and technical expertise.

The transitions depicted in Figure 4.15 illustrate various trends and pathways. Many involve moving within cognitive and technical roles as workers seek positions that utilise their analytical and problem-solving skills in more sustainable occupations. Although transitions within physical roles are less frequent, some occur where workers move to positions with reduced automation risk, mainly due to the high risk of automation in occupations where physical tasks predominate, as mentioned above. For example, transitions to "Travel Attendants and Travel Stewards" include movements from "Bartenders", "Cleaning", and "Housekeeping Supervisors in offices, hotels and other establishments", "Domestic Housekeepers", and "Transport Conductors", emphasising the adaptability of service-oriented skills.

To conclude, the analysis of expected occupation transitions based on skill similarity underscores the importance of strategic career planning and skill development. By leveraging the insights from Figure 4.15, workers, employers, and policymakers can better navigate the evolving labour market, ensuring they are prepared for future challenges. This integrated view of skills and occupations through cluster analysis provides an excellent understanding of the competency landscape, essential to individual career planning and development. By understanding these clusters and their characteristics, we can adjust training programmes to better align with the nature of the tasks involved, thereby increasing overall efficiency and effectiveness in various professional contexts.

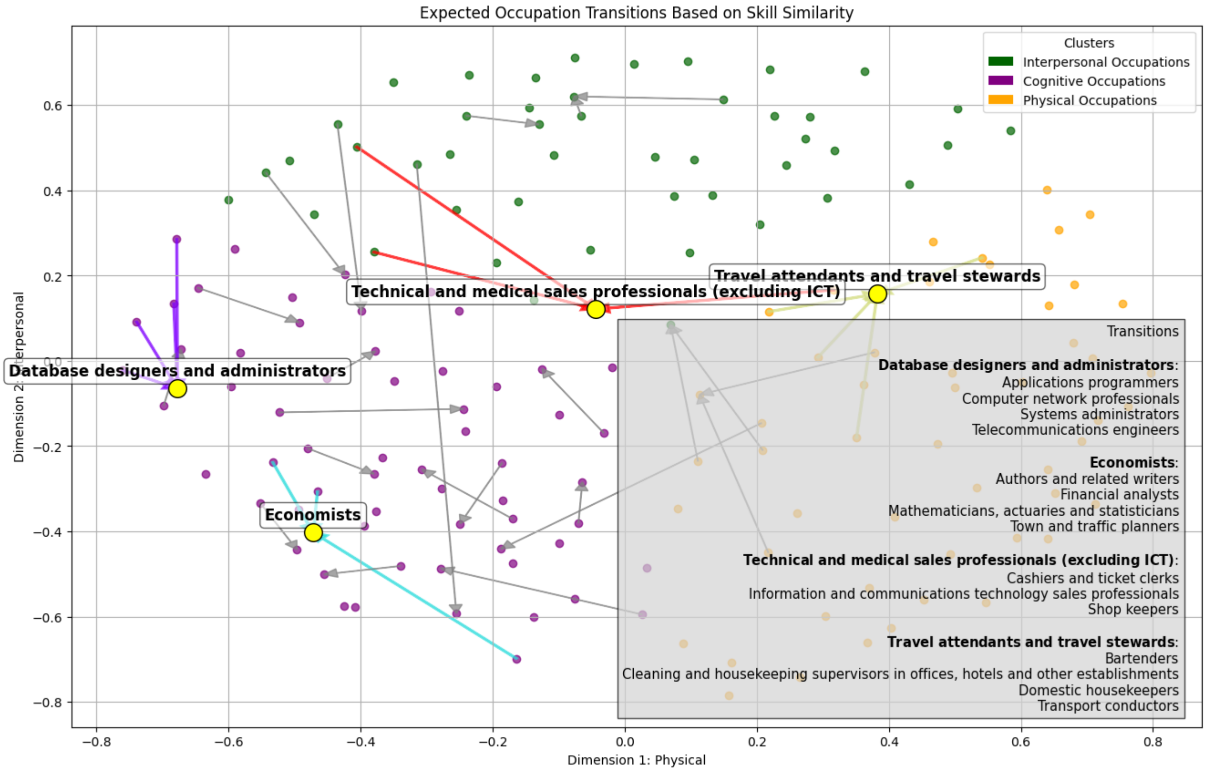

Figure 4.15 – Expected Occupation Transitions Based on Skill Similarity

4.5.2. Optimal Transitions

In this section, we examine the optimal transitions between occupations based on high skill similarity, focusing on pathways where workers can move from occupations with high automation risk to those with lower risk. Figure 4.16 provides a visual representation of these transitions. The key criteria for these transitions are a high similarity score (50% or more), a lower risk for the target occupation (less than 50%), and a higher risk for the base occupation (50% or more). This analysis identifies the best pathways with the greatest skill overlap, ensuring smoother worker transitions.

In total, 10 transitions meet these criteria. The similarity scores for these transitions range from approximately 51% to 66%, with an average similarity score of 54%, indicating a high degree of skill overlap. This high similarity suggests that workers can transfer their existing skills more easily, minimising the need for extensive retraining.

Figure 4.16 illustrates these transitions, highlighting several key examples. “Biologists, botanists, zoologists, and related professionals” show a transition from “chemical engineers”, reflecting the similarity in scientific principles and analytical skills. The transition score underscores the significant overlap in competencies required for these roles. “Database designers and administrators” transitioning from “systems administrators” demonstrate a strong foundational knowledge of IT principles and database management skills shared between these occupations. “Early childhood educators” display optimal transitions from “teachers' aides”, with similarity scores reflecting shared skills in child development, instructional strategies, and classroom management. “Economists” show optimal transitions from “mathematicians”, “actuaries and statisticians”, and “town and traffic planners”, highlighting shared skills in data analysis, economic modelling, and quantitative reasoning. “Electronics engineers” transitioning from “systems analysts” and “industrial and production engineers” reflect commonalities in engineering principles, problem-solving, and technical expertise across these disciplines. “Mechanical engineers” transitioning from various engineering roles, such as “industrial and production engineers”, underscore the versatility of engineering skills. “Philosophers, historians, and political scientists” transitioning from “translators, interpreters, and other linguists” highlight shared competencies in language, critical thinking, and research skills. “Physicists and astronomers” transitioning from “meteorologists” demonstrate optimal transitions based on shared skills in data analysis, scientific research, and quantitative reasoning. “Psychologists” transitioning from “pharmacists” underscore the overlap in patient care, counselling, and understanding human behaviour. “Web and multimedia developers” transitioning from “software developers” highlight shared competencies in coding, design, and technology use.

These optimal transitions emphasise the importance of identifying pathways where workers' existing skills can be utilised effectively in new roles with lower automation risk. By focusing on high-similarity transitions, we can ensure that workers face fewer barriers when moving to more secure occupations. This approach supports individual career development and helps in workforce planning by identifying and promoting viable career pathways within industries.

In conclusion, analysing optimal transitions based on skill similarity provides valuable insights into how workers can move from high-risk to lower-risk occupations with minimal retraining. The high overlap of skills in these transitions ensures smoother career shifts, enhancing job security. Figure 4.16 visually encapsulates these transitions, offering a clear roadmap for workers and policymakers to navigate the challenges raised by automation and technological advancements. By leveraging these insights, we can promote a more adaptable and secure workforce, prepared for the evolving demands of the labour market.

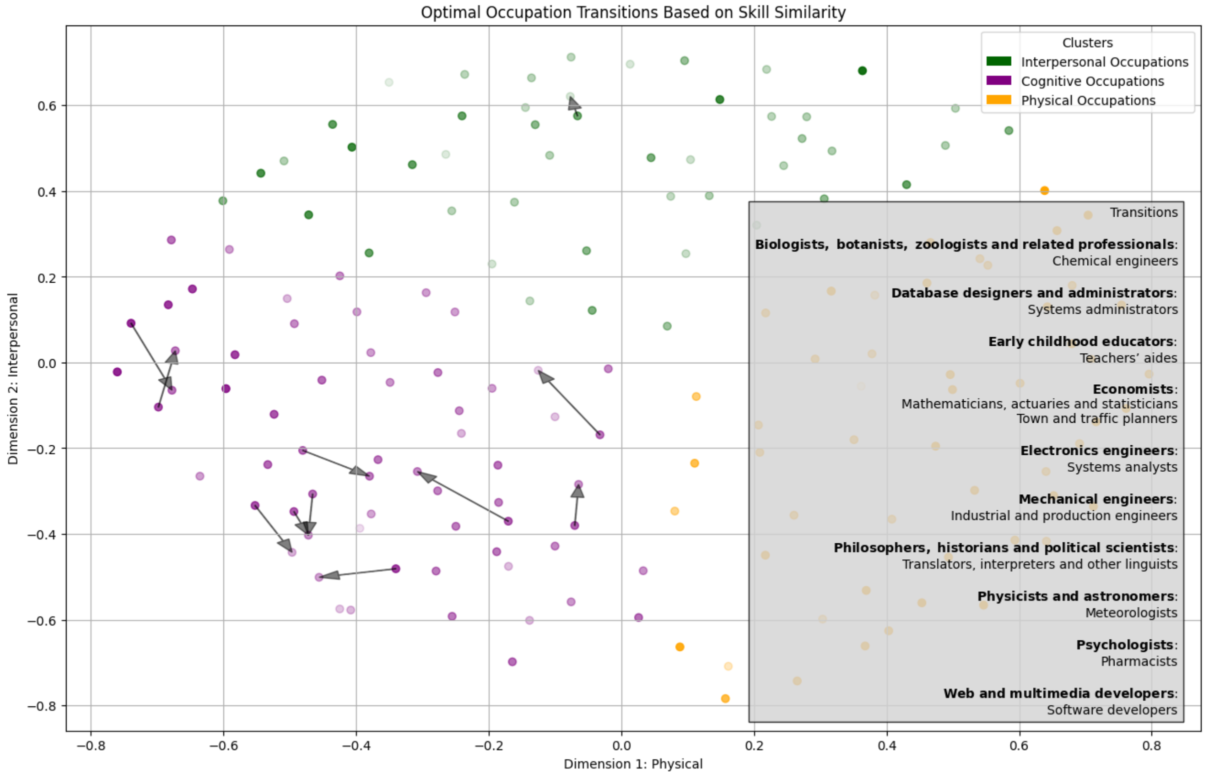

Figure 4.16 – Optimal Occupation Transitions Based on Skill Similarity

4.6. Web Application

To transform all this information into something valuable and tangible, a web application was developed to allow the target audience of this project to interact and benefit from its features. The interface was designed with simplicity as a core value, ensuring that all Generation X professionals feel comfortable and confident using it without feeling overwhelmed by complexity.

When users start the application, they are greeted by the homepage, styled as a dashboard, as shown in Appendix E, Figure E1. Users can search for an occupation they wish to analyse, with similar occupations suggested as they type in the search bar. Additionally, the homepage displays two histograms of the five occupations with the highest and lowest risk of automation.

Upon searching for a specific occupation, a risk meter is displayed, indicating the automation risk associated with that occupation, as demonstrated in Appendix E, Figure E2. Users then have access to two key functionalities for the analysed occupation. First, they can view the criteria used to calculate the automation risk, including the tasks involved, their importance, their frequency, and their automation risk probability. Users can also personalise the structure of the occupation by adding or removing tasks and defining the importance and frequency of each task, which updates the automation risk score automatically. This feature is illustrated in Appendix E, Figure E3.

One of the most compelling features is the skills and competences guide for a smooth and successful transition. Based on the searched occupation, the application presents the five most similar occupations in terms of skills that also have a lower automation risk. Users can further filter these suggestions using scales to set preferred ranges for automation risk score and similarity to the current occupation. Additionally, there is a search bar to choose a specific occupation for transition, providing users with even more control and flexibility. An interesting feature is the preset button for optimal transitions, which only shows job suggestions with an automation risk below 50% (low-risk categories). Although throughout this project, we analysed optimal transitions with a similarity of 50% or more, in addition to an automation risk below 50%, in the application, we did not strictly respect this criterion to avoid the absence of suggestions for some professions, thus adapting the term "optimal" for reasons of practical applicability. All these functionalities are illustrated in Appendix E, Figure E4.

As represented in Appendix E, Figure E5, after selecting a target occupation, users can access the skills and competences guide, which is organised into three distinct tables, as previously defined.

With all this information presented in a structured and accessible manner, users can make well-informed decisions, ensuring a smooth and confident transition to a new occupation. This thorough planning enhances the likelihood of a successful career change, empowering users with the knowledge and tools they need.

5. CONCLUSIONS

This research explored the potential impacts of AI and automation on occupations for Generation X professionals in Portugal, focusing on identifying expected and optimal career transitions. Utilising the CRISP-DM framework, innovative methodologies provided valuable insights for workforce planning and career development. By fine-tuning a pre-trained BERT model to predict the importance and frequency of tasks within occupations using ESCO occupation names, descriptions, and O*NET task descriptions, this research significantly enhanced the ESCO database, aligning it with the detailed level of O*NET. The model achieved an impressive accuracy of 90.03%, demonstrating the validity and reliability of the upgraded ESCO for future studies. This enhancement makes the ESCO database more comprehensive and useful for future European studies, especially with the rapid evolution of AI already outperforming humans in some capabilities (Giattino et al., 2023).

Combining the upgraded ESCO dataset with a multimodal language model, GPT-4, we assessed the automation risk of the three largest occupation groups among Generation X in Portugal. The findings revealed that, as stated by previous studies, elementary occupations face the highest risks of automation due to their repetitive and low-skilled tasks in a physical environment (Brynjolfsson & McAfee, 2011, 2014; Nedelkoska & Quintini, 2018; Xu et al., 2023). This situation presents a concerning panorama for individuals in these roles, as both optimal and valid transitions to lower-risk occupations are non-existent, suggesting an urgent need for significant retraining and adaptation.

On the other hand, professionals and service and sales workers in more technical environments show lower, though not negligible, susceptibility to automation. Detailed task-level analysis indicates that even high-skilled tasks involving complex problem-solving are at risk, as evidenced by occupations like “Application programmers” and “Computer network professionals.” This scenario suggests that automation is spreading to more highly qualified occupations, indicating that over time, the complexity of activities or the environment of occupation is becoming less relevant to the automation risk they present. We can, therefore, conclude that tasks with well-established rules, regardless of complexity, can be automated (Lassébie and Quintini, 2022; Giattino et al., 2023; Gmyrek et al., 2023).

On a more positive note, occupations requiring interpersonal skills, such as communication, collaboration, and creativity, are currently the safest. These roles pose a significant challenge for AI replication, providing a potential area of job security for Generation X professionals.

To support Generation X professionals in Portugal, we developed an interactive web application named Career Xplorer. This tool integrates all the information gathered into a recommendation system, providing personalised career transition guidance and a skills and competence guide to ensure smoother and more confident transitions, enhancing job security, workforce resilience, and even promoting greater job happiness. This application is not only a practical tool for workers but also crucial for educators, and policymakers. By focusing on reskilling and upskilling initiatives that respond to the specific needs of different professional groups, we are promoting large-scale preparation of the workforce for the changes related to automation and technological evolution. In this way, individual treatment reduces the impact of automation, allowing the worker to continue adapting to a new type of occupation with job stability, thus promoting higher job satisfaction rates.

Overall, this project significantly contributes to the field by offering a new, data-driven approach to understanding and mitigating the impacts of AI and automation on the workforce. It provides a practical framework for career transition planning, supporting Generation X professionals in Portugal and potentially serving as a model for similar initiatives in other regions. The integration of advanced AI models and comprehensive databases has paved the way for more informed and strategic career planning, ensuring that workers are better prepared to navigate the challenges and opportunities presented by an AI-driven future.

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

One of the primary limitations of this study is its focus on the three most frequent 1-digit occupation groups within Generation X in Portugal—Professionals, Service and Sales Workers, and Elementary Occupations. This selective focus, necessitated by computational limitations, means our findings do not apply to all occupation groups and may not be easily comparable with those of similar studies. However, our methodology introduces an innovative approach to providing personalised occupational data, enhancing the understanding of occupational similarities, skill requirements, and automation risks.

Another limitation was the restricted hyperparameter tuning due to computational costs. The complexity of the model made extensive local runs impractical, potentially limiting optimisation. Despite these constraints, our model achieved good performance, indicating that this limitation did not severely impact the overall results.

The cost of using the GPT-4 API for calculating automation risk for each task also posed a limitation. This study calculated risks for numerous tasks, adding to the overall expense. Furthermore, validation of automation risk by a group of specialists would have provided a closer approximation to real-world scenarios, reducing potential bias and enhancing the applicability of our findings.

Additionally, this study, like many others, overlooks the potential for new job creation due to automation. As automation decreases output prices and increases demand, new jobs may emerge in existing and new fields. Therefore, our analysis might be more alarming than the actual future, especially if there is an improvement in skills.

Future research should expand the scope to include all 1-digit occupation groups, providing a more comprehensive analysis and extending applicability to a wider range of occupations and generational groups. Incorporating additional metrics such as salary, academic requirements, job offers, course availability, and location could offer a more holistic view of the occupational landscape and enhance the recommendation system's relevance at the European level.

User feedback will be crucial for future adaptations, ensuring the system remains relevant and user-friendly. This feedback can guide iterative improvements in functionality and interface design. Launching the platform online will maximise its impact and accessibility, allowing real-time interaction for dynamic career planning and management. Integrating APIs for real-time data from sources like ESCO and O*NET, as well as job and course offers, will enhance the system's responsiveness and accuracy.

By addressing these recommendations, future research can improve the robustness and applicability of the findings, providing comprehensive support for individuals navigating the evolving labour market due to AI and automation advancements.

BIBLIOGRAPHICAL REFERENCES

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Labor Economics, 40(S1), S293-S340. doi: https://doi.org/10.1086/718327

Arntz, M., Gregory, T., & Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. OECD Social, Employment and Migration Working Papers, No. 189. OECD Publishing, Paris. https://doi.org/10.1787/5jlz9h56dvq7-en

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in Organizations: Current State and Future Opportunities. MIS Quarterly Executive, 19(4), Article 4. Available at SSRN: https://ssrn.com/abstract=3741983 or http://dx.doi.org/10.2139/ssrn.3741983

Brattberg, E., Rugova, V., & Csernatoni, R. (2020). Europe and AI: Leading, lagging behind, or carving its own way? (Vol. 9). Washington, DC, USA: Carnegie Endowment for International Peace.

Brynjolfsson, E., & McAfee, A. (2011). Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Brynjolfsson and McAfee, Lexington, MA.

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

Burrus, J., Jackson, T., Xi, N., & Steinberg, J. (2013). Identifying the most important 21st century workforce competencies: An analysis of the Occupational Information Network (O* NET). ETS Research Report Series, 2013(2), i-55.

Chapman, P. (2000). CRISP-DM 1.0: Step-by-step data mining guide.

Crouter, A. C., Lanza, S. T., Pirretti, A., Goodman, W. B., Neebe, E., & Family Life Project Key Investigators. (2006). The O* Net jobs classification system: A primer for family researchers. Family Relations, 55(4), 461-472.

Dave, V. S., Zhang, B., Al Hasan, M., AlJadda, K., & Korayem, M. (2018). A Combined Representation Learning Approach for Better Job and Skill Recommendation. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18). Association for Computing Machinery, New York, NY, USA, 1997–2005. https://doi.org/10.1145/3269206.3272023

Dawson, N., Williams, M. A., & Rizoiu, M. A. (2021). Skill-driven recommendations for job transition pathways. PLOS ONE, 16(8), e0254722. https://doi.org/10.1371/journal.pone.0254722

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. Pew Research Center, 17(1), 1-7.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv. http://arxiv.org/abs/2303.10130

European Commission. (2022). ONET ESCO Technical Report. esco.ec.europa.eu.

ESCO. (2024). Escopedia. Retrieved from https://esco.ec.europa.eu/en/about-esco/escopedia/escopedia

Fishman, A. A. (2016). How generational differences will impact America's aging workforce: strategies for dealing with aging Millennials, Generation X, and Baby Boomers. Strategic HR Review, 15(6), 250-257. https://doi.org/10.1108/SHR-08-2016-0068

Fonseca, T., Lima, F., & Pereira, S. C. (2018). Job polarization, technological change and routinization: Evidence for Portugal. Labour Economics, 51, 317-339. https://doi.org/10.1016/j.labeco.2018.02.003

Frank, M., Autor, D., Bessen, J., Brynjolfsson, E., Cebrian, M., Deming, D., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. Proceedings of the National Academy of Sciences of the United States of America, 116, 6531–6539.

Frey, C. B., & Osborne, M. A. (2013). The future of employment: How susceptible are jobs to computerization? University of Oxford, Oxford, England.

Georg Graetz, G., & Michaels, G. (2018). Robots at Work. The Review of Economics and Statistics, 100(5), 753–768. https://doi.org/10.1162/rest_a_00754

Giattino, C., Mathieu, E., Samborska, V., & Roser, M. (2023). Artificial Intelligence. OurWorldInData.org. Retrieved from https://ourworldindata.org/artificial-intelligence

Gmyrek, P., Berg, J., & Bescond, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality [ILO Working Paper No. 96]. International Labour Organization (ILO). https://doi.org/10.54394/FHEM8239

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review, 61(4), 5-14. https://doi.org/10.1177/0008125619864925

Handel, M. (2012). Trends in Job Skill Demands in OECD Countries. OECD Social, Employment and Migration Working Papers, No. 143. OECD Publishing, Paris. https://doi.org/10.1787/5k8zk8pcq6td-en

Hilton, M. L., & Tippins, N. T. (Eds.). (2010). A database for a changing economy: Review of the Occupational Information Network (O* NET).

International Labour Organization. (2012). International Standard Classification of Occupations: ISCO-08. Retrieved from https://www.ilo.org/public/english/bureau/stat/isco/

Jenkins, R., & Moses, E. (2014). The impact of O*NET on career guidance. Career Development Quarterly, 62(3), 234-245.

Jorgensen, B. (2003). Baby Boomers, Generation X and Generation Y? Policy implications for defence forces in the modern era. Foresight, 5(4), 41-49. https://doi.org/10.1108/14636680310494753

Karger, E., Rosenberg, J., Jacobs, Z., Hickman, M., Hadshar, R., Gamin, K., Smith, T., Williams, B., McCaslin, T., & Tetlock, P. E. (2023). Forecasting existential risks: Evidence from a long-run forecasting tournament [FRI Working Paper #1]. Forecasting Research Institute.

Lassébie, J., & Quintini, G. (2022). What skills and abilities can automation technologies replicate and what does it mean for workers?: New evidence. OECD Social, Employment and Migration Working Papers, No. 282. OECD Publishing, Paris. https://doi.org/10.1787/646aad77-en

Lane, M., Williams, M., & Broecke, S. (2023). The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers. OECD Social, Employment and Migration Working Papers, No. 288. OECD Publishing, Paris. https://doi.org/10.1787/ea0a0fe1-en

Małkowska, A., Urbaniec, M., & Kosała, M. (2021). The impact of digital transformation on European countries: insights from a comparative analysis. Equilibrium. Quarterly Journal of Economics and Economic Policy, 16(2), 325–355. doi: https://doi.org/10.24136/eq.2021.012

Martin, J. (2018). Skills for the 21st century: Findings and policy lessons from the OECD survey of adult skills. OECD Education Working Papers, No. 166. OECD Publishing, Paris. https://doi.org/10.1787/96e69229-en

M. M. Nair, Tyagi, A. K., & Sreenath, N. (2021). The Future with Industry 4.0 at the Core of Society 5.0: Open Issues, Future Opportunities and Challenges. 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2021, pp. 1-7. https://doi.org/10.1109/ICCCI50826.2021.9402498

National Center for ONET Development. The ONET® Content Model. O*NET Resource Center. Retrieved from https://www.onetcenter.org/content.html

Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training. OECD Social, Employment and Migration Working Papers, No. 202. OECD Publishing, Paris. https://doi.org/10.1787/2e2f4eea-en

Organisation for Economic Co-operation and Development (OECD). (2024). OECD Data. Retrieved May 28, 2024, from https://www.oecd.org/about/

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleishman, E. A. (2001). Understanding Work Using the Occupational Information Network (O*NET): Implications for Practice and Research. Personnel Psychology, 54(2), 451-492.

Plumed, F., Contreras Ochando, L., Ferri, C., Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M., & Flach, P. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 1-1. https://doi.org/10.1109/TKDE.2019.2962680

PORDATA. (2022). Número de divórcios por 100 casamentos (%) (INE, INE | DGPJ/MJ). Retrieved from https://www.pordata.pt/portugal/numero+de+divorcios+por+100+casamentos-531

Smedt, J.D., Vrang, M.L., & Papantoniou, A. (2015). ESCO: Towards a Semantic Web for the European Labor Market. LDOW@WWW.

Steiner, A.M., & Fletcher, P.C. (2017). Sandwich Generation Caregiving: A Complex and Dynamic Role. Journal of Adult Development, 24, 133–143. https://doi.org/10.1007/s10804-016-9252-7

Tai, M. C.-T. (2020). The impact of artificial intelligence on human society and bioethics. Tzu Chi Medical Journal, 32(4), 339-343. https://doi.org/10.4103/tcmj.tcmj_71_20

Tavakoli, M., Mol, S., & Kismihók, G. (2020). Labour Market Information Driven, Personalized, OER Recommendation System for Lifelong Learners. In Proceedings of the 12th International Conference on Computer Supported Education - Volume 2: CSEDU; ISBN 978-989-758-417-6; ISSN 2184-5026, SciTePress, pages 96-104. https://doi.org/10.5220/0009420300960104

World Economic Forum. (2023). The Future of Jobs Report 2023. https://www.weforum.org/publications/the-future-of-jobs-report-2023/

World Economic Forum. (2024). Global Risks Report 2024. https://www.weforum.org/publications/global-risks-report-2024/

Xu, D., Yang, H., Rizoiu, M.-A., & Xu, G. (2023). From Occupations to Tasks: A New Perspective on Automatability Prediction Using BERT, 2023 10th International Conference on Behavioural and Social Computing (BESC), Larnaca, Cyprus, 2023, pp. 1-7. https://doi.org/10.1109/BESC59560.2023.10386407

APPENDIX A

SUPPORTING DATA TABLES AND STRUCTURES

	ESCO/ISCO Code	ESCO/ISCO Title	O*NET-SOC 2019 Code	O*NET-SOC 2019 Title
0	0110.10	lieutenant	33-1011.00	First-Line Supervisors of Correctional Officers
1	0110.10	lieutenant	33-1012.00	First-Line Supervisors of Police and Detectives
2	0110.10	lieutenant	33-3012.00	Correctional Officers and Jailers
3	0110.10	lieutenant	33-3021.00	Detectives and Criminal Investigators
4	0110.10	lieutenant	33-3051.00	Police and Sheriff's Patrol Officers

Figure A1 – Crosswalk Dataframe

	iscoCode	occupationName	occupationDescription	task
0	2111	Physicists and astronomers	Physicists and astronomers conduct research an...	conducting research and improving or developin...
1	2111	Physicists and astronomers	Physicists and astronomers conduct research an...	conducting experiments, tests and analyses on ...
2	2111	Physicists and astronomers	Physicists and astronomers conduct research an...	evaluating results of investigations and exper...
3	2111	Physicists and astronomers	Physicists and astronomers conduct research an...	applying principles, techniques and processes ...
4	2111	Physicists and astronomers	Physicists and astronomers conduct research an...	ensuring the safe and effective delivery of ra...

Figure A2 – ESCO Task Dataframe

	iscoCode	ISCOgroupName	skillName	skillDescription	skillCode	relationType
0	2654	Professionals	planning and scheduling events and activities	Planning and scheduling workflow and timelines...	S4.2.2	essential
1	2654	Professionals	performing risk analysis and management	Identify and assess factors that may have nega...	S2.7.5	essential
2	2654	Professionals	coordinating activities with others	Communicating and liaising with colleagues, cl...	S1.2.1	essential
4	2654	Professionals	negotiating and managing contracts and agreements	Negotiating and managing contracts and agreeme...	S1.1.1	essential
5	2654	Professionals	advising on workplace health and safety issues	Promoting and providing information and advice...	S1.5.9	essential
...	...	...	...	...	...	...
38491	2422	Professionals	providing information to the public and clients	Answering questions, making recommendations, a...	S3.4.1	essential
38495	2422	Professionals	purchasing goods or services	Buying and ordering goods, services and supplies.	S1.6.3	optional
38500	2422	Professionals	recruiting and hiring	Acquiring the right talent to achieve the orga...	S4.7.0	optional
38501	2422	Professionals	monitoring safety or security	Monitoring and inspecting the safety or securi...	S2.8.4	optional
38509	2422	Professionals	executing financial transactions	Executing financial or commercial transactions...	S4.4.2	optional

Figure A3 – ESCO Skills Datafame

	O*NET-SOC Code	Title	Description
0	11-1011.00	Chief Executives	Determine and formulate policies and provide o...
1	11-1011.03	Chief Sustainability Officers	Communicate and coordinate with management, sh...
2	11-1021.00	General and Operations Managers	Plan, direct, or coordinate the operations of ...
3	11-1031.00	Legislators	Develop, introduce, or enact laws and statutes...
4	11-2011.00	Advertising and Promotions Managers	Plan, direct, or coordinate advertising polici...
...	...	...	...
1011	55-3014.00	Artillery and Missile Crew Members	Target, fire, and maintain weapons used to des...
1012	55-3015.00	Command and Control Center Specialists	Operate and monitor communications, detection,...
1013	55-3016.00	Infantry	Operate weapons and equipment in ground combat...
1014	55-3018.00	Special Forces	Implement unconventional operations by air, la...
1015	55-3019.00	Military Enlisted Tactical Operations and Air/...	All military enlisted tactical operations and ...

Figure A4 – O*NET Occupation Dataframe

	onetOccupationCode	occupationName	occupationDescription	taskCode	task	scaleValue	frequencyDescription	original
0	25-1021.00	Computer Science Teachers, Postsecondary	Teach courses in computer science. May special...	5697	Participate in student recruitment, registrati...	1	Yearly or more	True
1	25-1041.00	Agricultural Sciences Teachers, Postsecondary	Teach courses in the agricultural sciences. In...	5786	Plan, evaluate, and revise curricula, course c...	1	Yearly or more	True
2	27-3091.00	Interpreters and Translators	Interpret oral or sign language, or translate ...	9338	Discuss translation requirements with clients ...	1	Yearly or more	True
3	25-1066.00	Psychology Teachers, Postsecondary	Teach courses in psychology, such as child, cl...	6072	Act as advisers to student organizations.	1	Yearly or more	True
4	25-2058.00	Special Education Teachers, Secondary School	Teach academic, social, and life skills to sec...	6823	Sponsor extracurricular activities, such as cl...	1	Yearly or more	True

Figure A5 – Frequency Scale Dataframe

	onetOccupationCode	taskCode	task	scaleValue	frequencyDescription	original	escoOccupationCode	occupationName	occupationDescription	combined
0	11-3031.03	15666	Review offering documents or marketing materia...	1	Yearly or more	True	2412	Financial and investment advisers	Financial and investment advisers develop fina...	[OCC_NAME] Financial and investment advisers [...
1	11-3031.03	15660	Prepare for and respond to regulatory inquiries.	1	Yearly or more	True	2412	Financial and investment advisers	Financial and investment advisers develop fina...	[OCC_NAME] Financial and investment advisers [...
2	11-3031.03	15671	Meet with investors to determine investment go...	1	Yearly or more	True	2412	Financial and investment advisers	Financial and investment advisers develop fina...	[OCC_NAME] Financial and investment advisers [...
3	11-3031.03	15672	Identify group or individual target investors ...	1	Yearly or more	True	2412	Financial and investment advisers	Financial and investment advisers develop fina...	[OCC_NAME] Financial and investment advisers [...
4	11-3031.03	15662	Hire or evaluate staff.	1	Yearly or more	True	2412	Financial and investment advisers	Financial and investment advisers develop fina...	[OCC_NAME] Financial and investment advisers [...
...	...	...	...	...	...	...	...	...	...	...
10291	25-1082.00	6190	Engage in activities on campus and in the comm...	2	More than monthly to weekly	False	2310	University and higher education teachers	University and higher education teachers prepa...	[OCC_NAME] University and higher education tea...
10292	25-1121.00	6279	Join committees dealing with institutional pol...	2	More than monthly to weekly	False	2310	University and higher education teachers	University and higher education teachers prepa...	[OCC_NAME] University and higher education tea...
10293	25-1121.00	6279	Participate in committees that address institu...	2	More than monthly to weekly	False	2310	University and higher education teachers	University and higher education teachers prepa...	[OCC_NAME] University and higher education tea...
10294	25-1121.00	6279	Be a member of committees handling institution...	2	More than monthly to weekly	False	2310	University and higher education teachers	University and higher education teachers prepa...	[OCC_NAME] University and higher education tea...
10295	33-1021.00	22919	Perform maintenance and minor repairs on firef...	2	More than monthly to weekly	False	5411	Fire-fighters	Firefighters prevent, fight and extinguish fir...	[OCC_NAME] Fire-fighters [OCC_DESC] Firefighte...

10296 rows × 10 columns

Figure A6 – Training Dataframe

APPENDIX B

DETERMINING OPTIMAL CLUSTERS

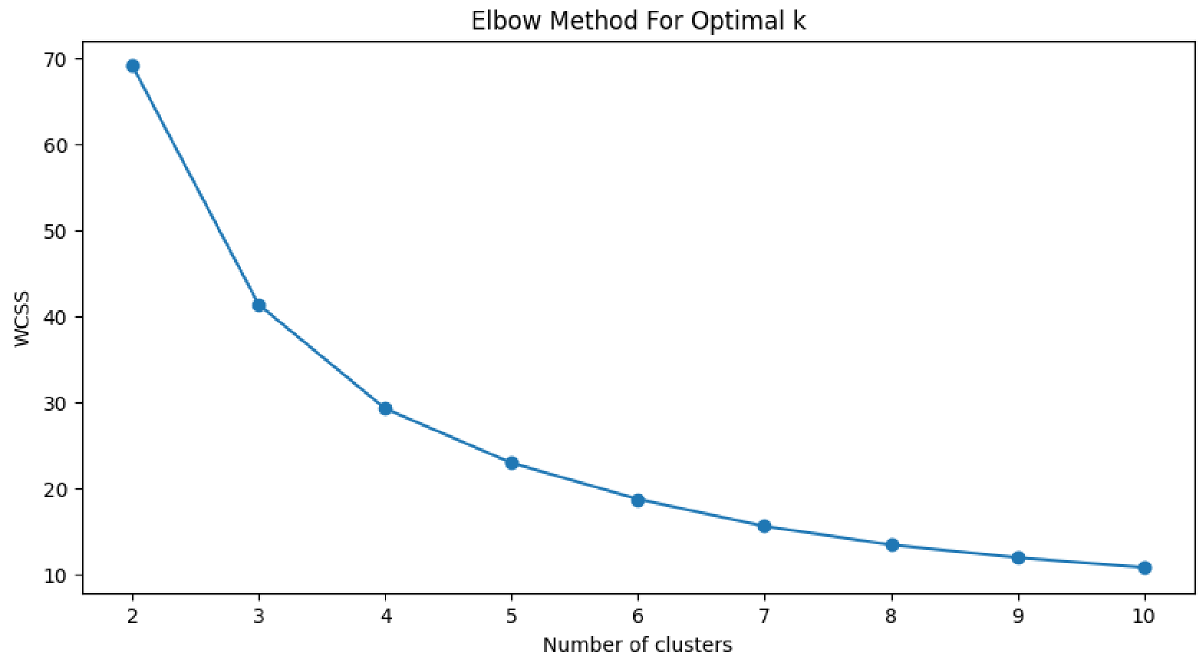

Figure B1 – Elbow Method for Optimal K

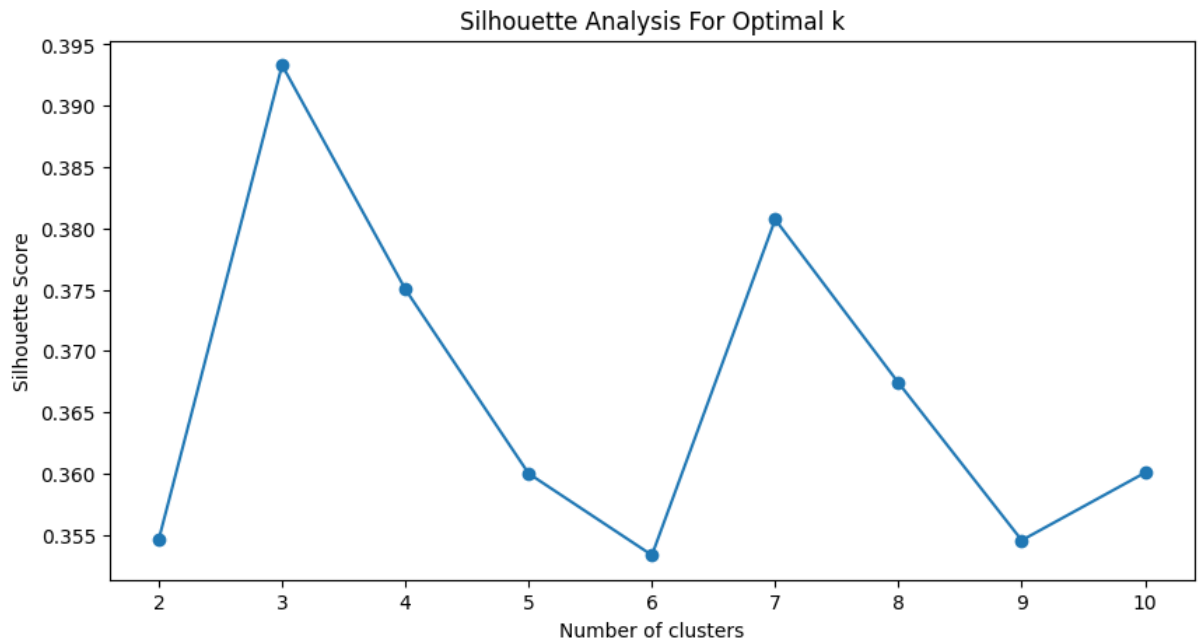

Figure B2 – Silhouette Analysis for Optimal K

APPENDIX C

Model Performance Metrics

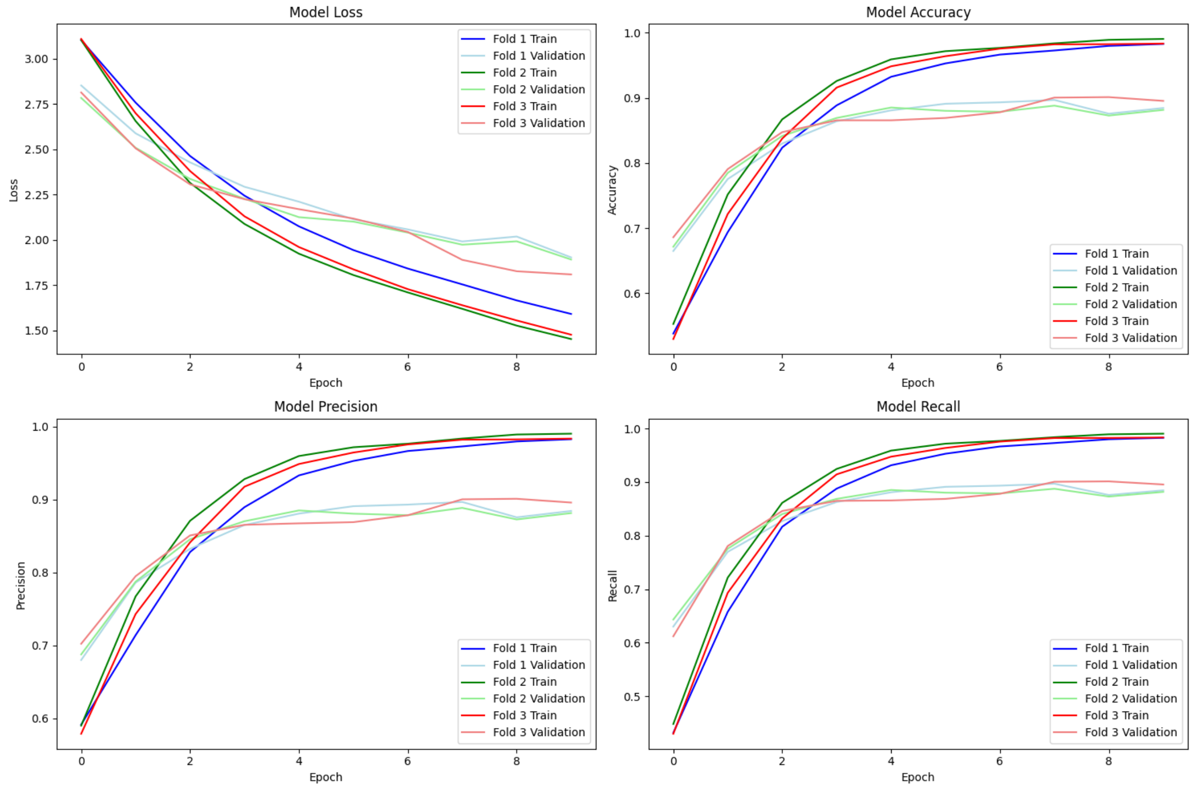

Figure C1 – Performance Metrics for Frequency Prediction

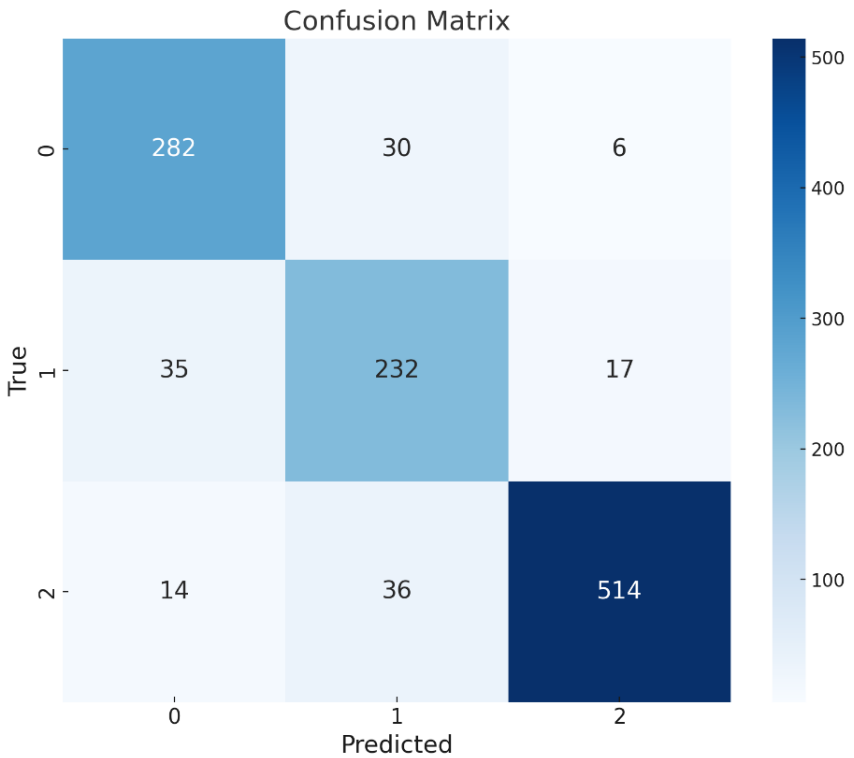

Figure C2 – Confusion Matrix of Frequency Prediction

APPENDIX D

Distribution of Task Frequency Classes

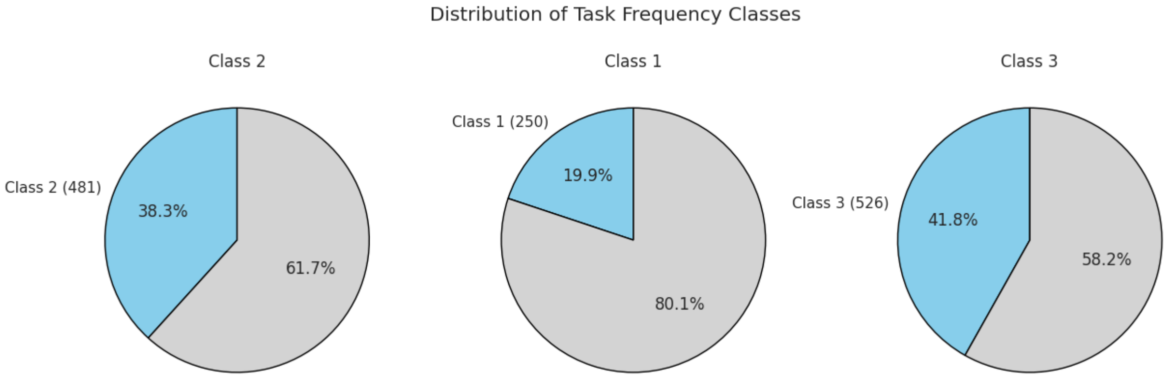

Figure D1 – Distribution of Task Frequency Classes

APPENDIX E

CAREER XPLORER INTERFACE

Figure E1 – Homepage from Web Application

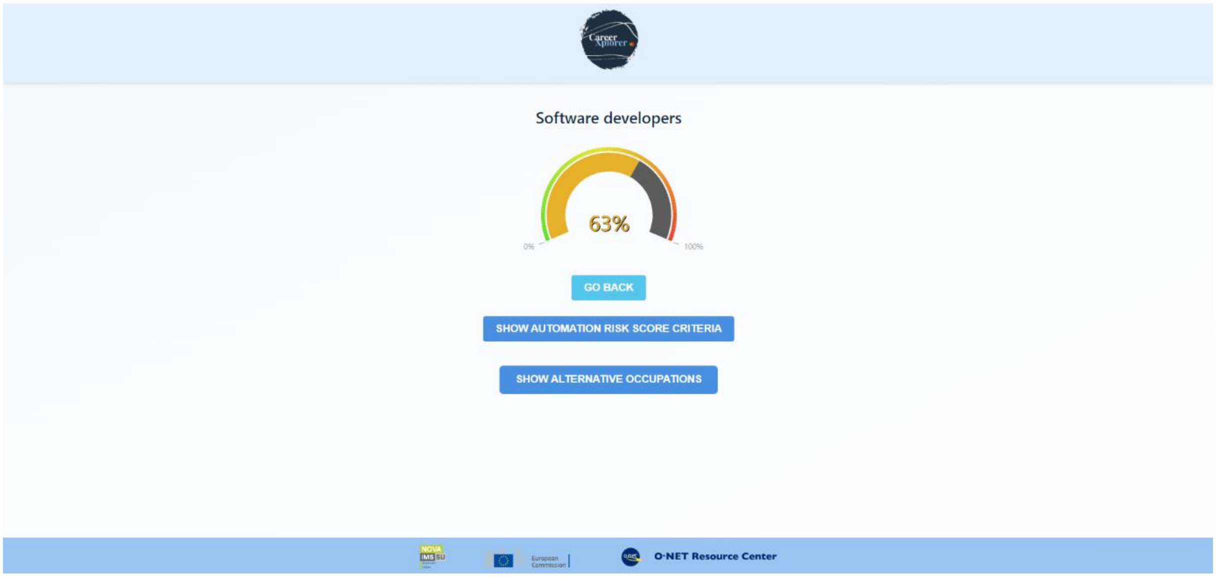

Figure E2 – Risk Automation Score Meter from Web Application

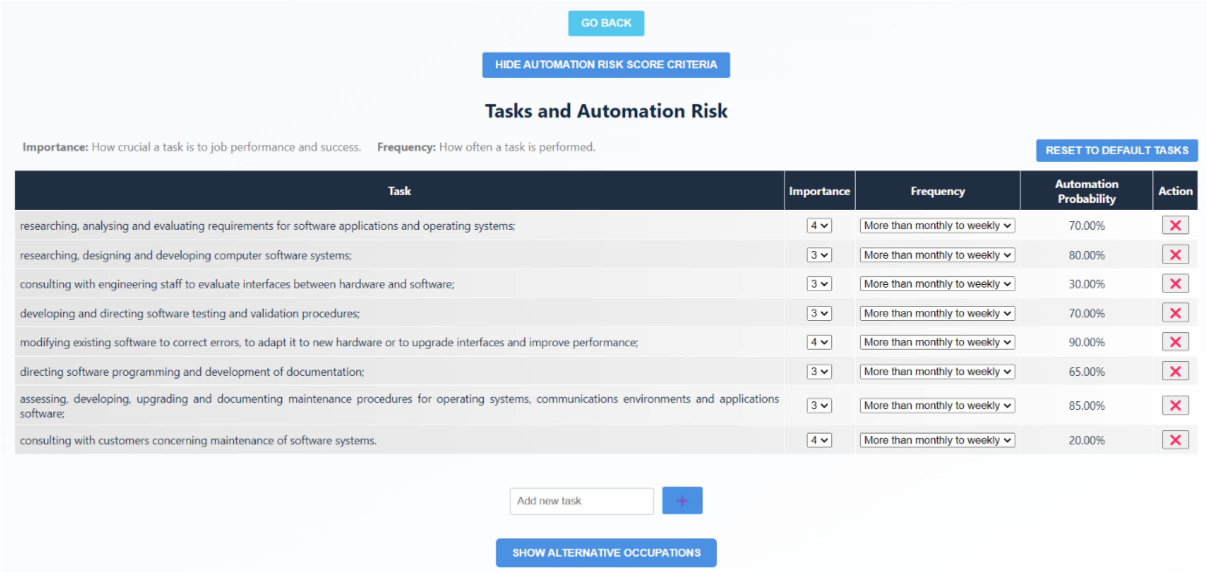

Figure E3 – Automation Risk Score Criteria from Web Application

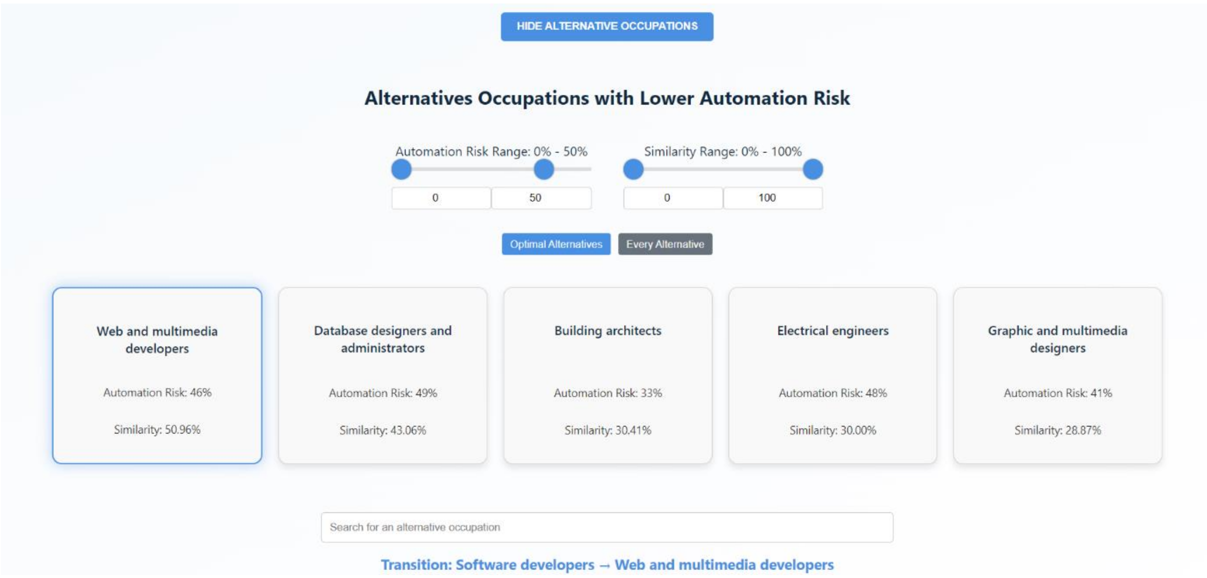

Figure E4 – Alternative Occupations from Web Application

Transition: Software developers → Web and multimedia developers

High Importance & High Distance Skills

Skills with high importance in the target occupation and significant differences from the source occupation.

Seize the opportunity to grow and excel in these critical areas!

Skill	Skill Description
artistic and creative writing	Writing original text of a creative nature for artistic, journalistic, promotional or similar purposes.
browsing, searching and filtering digital data	Articulating information needs, searching for data, information and content in digital environments, accessing and navigating them. Creating and updating personal search strategies.
developing financial, business or marketing plans	Envision and develop strategies and plans for companies and organisations aimed at purposes such as establishing new markets, refurbishing the equipment and machinery of a company, implementing pricing strategies, and restoring or maintaining financial viability.
monitoring and evaluating the performance of individuals	Monitoring the behaviour or performance of workers, students or oneself to ensure that work is completed satisfactorily and to evaluate performance, capabilities and training needs.
performing calculations	Performing mathematical calculations on financial, spatial, scientific or other data with or without the used of electronic tools.
promoting products, services, or programs	Promoting, advertising and marketing goods, services, programs or policies.
using digital tools for collaboration and productivity	Using ICT software and hardware to collaborate and communicate with others and to improve productivity.
using digital tools for processing sound and images	Using ICT software and hardware to for processing sound and images.
writing and composing	Writing text or composing music, creating original work with regard to format, style and content.

Low Importance Skills

Skills with lower importance in both the source and target occupations, indicating areas of less focus.

Keep these on your radar for a well-rounded skill set, but prioritize the essentials!

Skill	Skill Description
analysing business operations	Analysing and evaluating information and data on production and business operations.
coordinating activities with others	Communicating and liaising with colleagues, clients and other agencies on operational matters, problems and activities. Cooperating and liaising with outside agencies, clients and other organisational units to adapt the timing and nature of the activities.
documenting and recording information	Maintaining records of information, transactions and activities in digital, paper or other forms.
implementing new procedures or processes	Implementing new business procedures or processes to resolve practical, operational or conceptual problems which arise in the execution of work in a wide range of contexts.
preparing financial documents, records, reports, or budgets	Preparing and maintaining records and standardized reports on transactions, sales, financial information and budgets.
protecting ict devices	Making use of tools and methods to protect and maximize security of ICT devices and information by controlling access, such as by requiring passwords, digital signatures, and biometric identification, and by protecting systems through the use of software.
technical or academic writing	Writing original text of a primarily functional, technical or academic nature; editing text.
translating and interpreting	Translate one language into another language in written or spoken form. Match words and expressions with their corresponding brothers in other languages, while making sure that the message and nuances of the original text are preserved.
using word processing, publishing and presentation software	Using ICT software and hardware to develop and edit the content of documents and audio-visual and prepare them for publication.

Low Distance Skills

Skills with minimal differences in importance between the source and target occupations, suggesting areas of compatibility.

Leverage these familiar skills as a solid foundation for your transition!

Skill	Skill Description
complying with legal and organisational guidelines	Ensuring compliance with rules, standards, policies, guidelines or laws relating to matters other than health, safety and the environment.
conducting investigations	Conducting studies, investigations, research or surveys to increase knowledge and understanding.
developing operational policies and procedures	Developing organisational and operational methods, policies, procedures or standards.
gathering information from physical or electronic sources	Collecting, compiling or gathering information from digital, paper or other sources.

O*NET Resource Center

Figure E5 – Transition Guide Tables from Web Application

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa