Prototyping Neural Network-Based NLP Solutions for Person-Skill Fit Assessment

Vincent Roelofs

Graduate School of Natural Sciences - Utrecht University

First supervisor: Prof. dr. ir. Hajo Reijers

Second supervisor: Prof. dr. Eva Knies

19 December

Acknowledgements

I would like to commence by expressing my profound gratitude to Prof. dr. ir. Hajo Reijers. I greatly admire his ability to maintain a clear overview of the most salient aspects of this thesis over an extended period, while still posing specific and insightful questions about topics we discussed months ago. His guidance has taught me a great deal about taking ownership of the research process and its progression. I would also like to thank my second supervisor, Prof. dr. Eva Knies, for providing an invaluable external perspective to my IT-focused study, for she ensured that the research extended beyond the development of a theoretical IT artefact and was evaluated from this perspective. I am additionally deeply grateful to my parents, sister, and friends for their unwavering support and encouragement throughout this process. Their support was instrumental in sharing my excitement when significant progress was made and in maintaining morale during periods of experimentation where results were not as expected. Finally, I would like to express my gratitude to the interviewees for generously sharing their valuable insights.

Abstract

Organizations are shifting from a job-based to a task-based organizational structure (Chalutz-Ben Gal, 2023). This increased task focus requires assessing who would be fit for a task directly by utilizing person-skill fit instead of more vocational perspectives such as person-job fit (Chalutz-Ben Gal, 2023). Allocating tasks to suitable employees currently takes a significant amount of time for human research professionals (Bouajaja & Dridi, 2017). This paper explores how natural language processing (NLP) models based on neural networks can support people within organizations in efficiently identifying the most suitable individuals for specific tasks. A prototype system was developed and tested on a synthetically generated dataset of resumes based on the O*NET framework (National Center for O*NET Development, 2024b), to automate the process of allocating tasks to candidates. This was tested by utilizing large language models (LLMS), which proved unsuitable to accurately assess large amounts of resumes within a short amount of time. Vector embeddings were also tested to rank resumes based on person-skill fit. A quantitative analysis has shown a strong correlation between the ranking and the ability to perform the tasks (ρ = -0.7505). Domain experts who tested the prototype expressed satisfaction with its ranking and user-friendly design, emphasizing its potential to streamline HR processes and enhance efficiency. However, the reliance on synthetic data, must be addressed to confirm usability in real-world scenarios.

Table of contents

1. Introduction

1.1 Problem statement

Contemporary approaches to work organization, such as flexible work arrangements and cross-functional teams, have led to a shift in focus from employees with a single job at a single employer to a more fluid task-based approach to work (Chalutz-Ben Gal, 2023). While traditionally tasks were allocated based on job titles, the prevalence of task-based work renders a vocational level of analysis unsuitable for task allocation in modern work context (Chalutz-Ben Gal, 2023).Task allocation is generally taken to mean the assignment of various tasks that need to be done to resources, instruments, or agents capable of doing them or the allocation of such resources to tasks, which is equivalent (Sheridan, 1988).

The person-job fit determines the similarity between an individual's personality, skills, and competencies and the vocational environment (Kristof, 1996). However, due to the more task-based focus of modern organizational contexts, a more direct analysis between the fit of the skills of an individual and the tasks is needed, which is what the person-skill fit proposes (Chalutz-Ben Gal, 2023). Person-job fit is namely too broad of a level of analysis for assessing fit with individuals tasks, as it takes into account competencies for various tasks, combined with other characteristics such as a match with the organizational culture. Person-task fit on the other hand focusses solely on characteristics that are directly related to that specific tasks, making it more suited for gig-based work and other modern work arrangements where traditional measures of fit are less relevant. Allocating tasks to suitable employees takes a significant amount of time for human research professionals (Bouajaja & Dridi, 2017). While there are measurement scales for specific person-skill fits, such as in the paper by Audrin et al. (2024), utilizing this for task allocation would mean that a scale would have to be made for each possible person-skill fit. The

*Occupational Information Network (*O*NET*) database contains over 19,000 occupation-specific task statements* (National Center for O*NET Development, 2024a). Since each person-skill combination would require a separate scale, this would amount to over 19,000 scales, demanding significant time and resources to develop. As such, there is a need for an approach that is able to assess person-skill fit while not being reliant on specific scales, to make assessing person-skill fit more scalable.

Previous studies have recommended the implementation of a technological platform for person–skill fit assessment, but to date, such a platform that does person-skill fit assessment based on computer analysis does not exist (Chalutz-Ben Gal, 2023). While computer-based person-job fit assessment platforms exist, these platforms do not focus on individual tasks and are often focused on hiring the right candidate instead of on task allocation. Furthermore, there currently is no large-scale dataset available that can be used for the purpose of developing a method for task allocation based on resumes. While there are related projects that have created datasets that are focussed on specific domains such as IT jobs, such as the one presented by Jiechieu & Tsopze (2021), there are no large-scale datasets based on real resumes that are not domain-specific. Furthermore, these datasets are often focussed on matching skills to jobs, instead of matching skills and tasks (Jiechieu & Tsopze, 2021). While LLMs have previously been used to generate large-scale synthetic datasets of resumes, such as in the work by Prince Thapak (2023), these datasets lack any indication of whether someone is able to perform a certain task. Moreover, as these datasets are fully generated by LLMs, it cannot be proven that the content is an accurate reflection of real-world resumes.

1.2 Research objective

The objective of this study is to develop a platform that enables individuals within organizations to efficiently identify those best suited to specific tasks. The focus is on automating part of the assessment of person-task fit using neural-network-based NLP models. NLP is a tract of Artificial Intelligence and Linguistics, devoted to making computers understand the statements or words written in human languages (Khurana et al., 2023). By leveraging resume data, the platform developed in this research is designed to support non-technical users in making informed task-allocation decisions through a user-friendly, web-based interface. This platform could be used to help organizations that are in the process of reallocating tasks to more easily make informed decisions about the task allocation. An example of this would be an international organization with thousands of employees, of which it is unfeasible for the HR department to know all the skills of each employee. The prototype could then help to make a shortlist of potential candidates, after which the HR professionals can make the final decision. This way the prototype can cut down on the total time it takes for HR to shortlist potential candidates for a task, as well as help in making more informed decisions by taking more people into account than was feasible manually. As such, the goal of the prototype is not to be used at runtime but when tasks have to be allocated or reallocated by HR professionals.

1.3 Research questions

The research was centred around a set of research questions, with the primary research question being: ‘How can NLP models based on neural networks support people within organizations in efficiently identifying the most suitable individuals for specific tasks?’

The main question was formulated to include all relevant technologies used within the research, namely: LLMs, named entity recognition (NER) models, language recognition models, and vector embedding models. At the same time, the question excludes certain technologies, such as techniques that are not based on neural networks and only focus on exact word matches.

Moreover, the term NLP indicates that the focus is exclusively on text, and not on other modalities such as video or audio. The main question further clarifies that the prototype is not intended to autonomously rearrange tasks, but to support people in efficiently finding a suitable person to perform a task. To comprehensively address this main research question, the research is further divided into two subsidiary questions:

Sub-Research Question One: In what ways can neural network-based NLP models be used to assess the person-skill fit?'

This first sub question was used to explore how existing neural network-based NLP models can be used to aid in assessing a person's skill fit. Utilizing neural network-based NLP models was quite broad in scope, as it entailed both the generation of the Synthetic dataset, and utilizing these models for person-skill fit assessment.

Sub-Research Question Two: 'In what ways can neural network-based NLP models be made efficiently usable for non-technical people within organizations?'

This second sub-question was used to explore how a graphical user interface can be made that makes it easier for non-technical people to utilize these models. Neural network-based NLP models often require technical knowledge in order to use them directly, and the aim was to abstract the working of these models in a way that no technical knowledge was needed by the user, in order for these models to support in person-skill fit assessment.

These sub-questions aim to guide research through exploring various facets of task allocation in organizational settings, addressing both the technological and human aspects.

1.4 Contributions

1.4.1 Scientific

This research examines the use of NLP models based on neural networks to support people within organizations in efficiently identifying the most suitable individuals for specific tasks. This addresses a research gap regarding the potential usage of these models, as these models have not been widely researched in terms of their use in human resource (HR) practices. A key contribution of this study lies in its interdisciplinary insights, integrating knowledge from diverse research areas like business process management, HR management, and technology and Information Technology-related studies. This cross-disciplinary approach offers a comprehensive understanding of how different facets of these disciplines can be combined to improve task allocation within organizations.

This study also contributes to the scientific discipline by generating a dataset that contains over sixteen thousand synthetically generated resumes. This dataset is based on information from the O*NET database. This database provides comprehensive and detailed information about jobs and is developed by the U.S. Department of Labour (National Center for O*NET Development, 2024b). By basing the synthetic resumes on the *O*NET* dataset, it can be guaranteed that the content of the resumes corresponds to what information may be realistic on real-world resumes. By linking each synthetic resume to a specific job in the O*NET database, this dataset enables researchers to explore various job-related links, including, but not limited to task-specific capabilities aligned with real-world job requirements. The dataset is licensed under a *CC BY-NC 4.0* licence, making it available for other researchers to advance related studies.

1.4.2 Practical

This study examines the potential for creating a platform that can support people within organizations in efficiently identifying the most suitable individuals for specific tasks. In order to accomplish this, various neural network-based NLP models are used to analyse unstructured data from resumes. This NLP-based analysis contains various steps. Vector embeddings with hybrid retrieval are used to find the similarity between a resume and a task description. This similarity is then used to create a ranking of resumes in order to find the resume that is most similar to the task. The assumption behind this is that a resume that is similar to the task, most likely belongs to someone who is able to perform the task.

NER and models for detecting the language of written text have been used to extract structured information from the resume. LLMs are used to summarize key information from the resumes. The outcome of the various analysis is combined and shown in a web-based interface, with the intention of making it feasible for non-technical people within organizations to utilize the outcome of this analysis. The platform is designed to be able to be run independently without reliance on major cloud providers. Furthermore, the platform is designed to work well in multilingual context in order to enable multinational organizations to use the platform.

The proposed platform for person-skill fit assessment provides a scalable approach, eliminating the necessity for specific measurement scales for each person-skill fit assessment. The platform can be employed in the same manner for all tasks, facilitating the implementation of person-skill fit assessment for task assignment in organizations.

1.5 Scope

Various constraints, such as those related to available time and hardware, limit the scope of the research. This research utilizes synthetically generated data that is meant to be a realistic representation of real-world resumes. It is outside the research scope to manually verify each resumes data; instead, the available information will be used under the assumption that it provides valuable insights. This choice was made due to the extensive time it would take to verify each resume manually. Implementing the technological platform for person–skill fit assessment in organizations to assess the impact of the reallocated tasks is not within the remit of this research, given the impracticality of expecting organizations to adhere to the recommended task assignments and see the impact of the reallocated task within the time that is available for this research. Most of the time was utilized to design the prototype, limiting the time left for real-world evaluation. It is also out of the scope of the research to extensively test all the functionality of the prototype. For example, the prototype has been designed in order to be able to assess the task-person fit of resumes in multiple languages. However, the synthetic resume dataset only contains English resumes, as they were directly based on the English O*NET dataset. While translating some of the resumes would have been possible, this would have meant having to verify that the translations are correct and still correspond to the original information in the O*NET database. Given the sheer number of resumes in the total dataset utilizing multilingual resumes was left out of scope.

Quantitative analysis is utilized to verify whether the person-skill fit assessment is correct. To verify the usability of the prototype two interviews with domain experts have been conducted. These interviews mainly focus on whether the prototype includes the necessary functionality to be usable. The interviews also include a small test to see whether domain experts agree with the person-skill fit ratings of the prototype, but this is kept limited, as the correctness of the person-skill fit is proven through statistical measurements. Since this research focuses on assessing person skill fit rather than extracting names or detecting languages verifying the correct extraction of names from resumes and the detection of languages is deemed out of scope for this research. The libraries that were used for this that already provide data about the average recall and precision of these models, the current research assumes those estimations to be correct. Thus, the evaluation focuses primarily on the most important part of the prototype, which is the ability to assess whether an individual is able to perform a certain task, based on resume information.

2. Background

This section provides background information on the current state of research into utilizing NLP models based on neural networks to support people within organizations in efficiently identifying the most suitable individuals for specific tasks. This background section is separate from the systematic literature review which will be used to research what existing solutions exist for using neural network-based NLP models for tasks allocation specifically, and how these solutions impact employees.

2.1 Dynamic capabilities

Traditionally, firms have analysed their competitive edge by emphasizing their exclusive resources and capabilities, such as patents and technical expertise, through the resource-based perspective (Barney, 1991). However, the recent emphasis on dynamic capabilities has somewhat altered this view by highlighting a firm's capacity to adapt to its changing environment as a crucial factor in maintaining a competitive advantage (C. L. Wang & Ahmed, 2007). Economic factors, such as growing digitalization and increased globalization, have increased competition in markets (Hatzichronoglou, 1996). This rendered the old theories for assessing a business's competitive advantage inadequate, as a more adaptable approach to doing business became important (Teece, 2007). As organizations place greater importance on rapidly responding to market events, due to an increased focus on dynamic capabilities, this also relates to how jobs are organized. This relates to structuring jobs more into tasks that can be reassigned, instead of a fixed combination of tasks for a specific function(Rogiers & Collings, 2024). The underlying premise is that job deconstruction into specific tasks allows for a more dynamic alignment and restructuring of needed skills for changing market situations (Weller et al., 2019).

In light of the increasing tendency for work to be structured around discrete tasks, there is a growing need for a task-focused approach. The prototype can facilitate this transition and assist with an increased focus on dynamic capabilities.

2.2 Person-environment fit

A crucial aspect of this study is to ascertain which tasks are most compatible with the skills of employees. This is closely related to the theory of person-environment fit, which comprises both supplementary and complementary fit. This research focusses on *complementary fit*, which concerns the extent to which an individual's talents align with the demands of their environment. In contrast, *supplementary fit* is concerned with the degree of similarity between individuals and their environment. (Muchinsky & Monahan, 1987) Further specific instantiations of person-environment fit can be identified, as fit may be a viable construct for various analysis levels, such as person-job fit (Goldstein & Smith, 1995). This is a relatively broad concept, as it determines the similarity between an individual's personality and the vocational environment. As such, it can be considered the most comprehensive level of analysis for examining an individual's fit with a work environment (Kristof, 1996).

Kristof (1996) suggests that the appropriate level of fit to use for research is primarily dependent on the a priori rationale for expecting fit to be relevant at a particular level. As the research presented in this paper is primarily concerned with reallocating tasks, rather than a more comprehensive vocational perspective, the concept of person-job fit appears to be too broad for this research. Chalutz-Ben Gal (2023) proposes that person-skill fit may be a more suitable level of analysis for current working practices. The prevalence of task-based work, such as freelancing, more dynamic task assignments within companies, and the increasing number of individuals holding multiple jobs, renders a vocational level of analysis unsuitable for this type of analysis (Chalutz-Ben Gal, 2023). One disadvantage of utilizing the person-skill level of analysis is that it is a novel approach, as it has only recently been proposed to better align work conditions such as remote work, which are relatively recent phenomena (Chalutz-Ben Gal, 2023). Consequently, it lacks the extensive existing research that can be drawn upon, in contrast to, for instance, person-job fit, which has a substantial body of previous research. Nevertheless, there is some degree of conceptual overlap between person–skill fit and person–job fit, which means that existing measurement scales can be used as inspiration (Chalutz-Ben Gal, 2023).

2.3 Large language models

LLMs are neural network-based models with billions of parameters that are trained for textual understanding and classification (Mandvikar, 2023). One of the first LLM that achieved state-of-the-art results on various NLP tasks was BERT (Devlin et al., 2018). While this model is still in use today, several other models have since been developed that have improved various aspects of the LLM. In this research the main LLM used is Mistral NeMo (Mistral AI team, 2024). This model distinguishes itself from other open-source models by having a sixteen times larger context length than similar sized models. The context length determines the number of words that can be given as input to the LLM. As such, using a LLM with a large context length allows using larger resumes as input. Furthermore, Mistral Nemo is specifically designed for multilingual understanding, which may be crucial for using the prototype in international organizations (Mistral AI team, 2024). Furthermore, Mistral Nemo has state-of-the-art world knowledge and reasoning capabilities, compared to similar sized LLMs (Mistral AI team, 2024). This is important for this research as both world knowledge and reasoning are needed for the ability to accurately assess person-skill fit. Another model that is multilingual and has a large context length is Qwen 2.5 by Qwen Team (2024), which is also used in this research as it has a version that is half the size of Mistral Nemo, making it less hardware intensive to run.

LLMs represent one of the ways that NLP models based on neural networks can be used to support people within organizations in efficiently identifying the most suitable individuals for specific tasks. As such, this research has used LLMs to assess person-task fit, as well as summarize resume information.

2.4 Vector embeddings

Vector embedding models are neural network-based models that represent text in a high-dimensional space, where texts with similar meaning are close together. The possibility of calculating distances between texts allows for the calculation of similarity between texts. In this research such similarity calculations are used to find the most similar resumes to task descriptions. One of the first papers which popularized the idea of utilizing word embeddings is Mikolov et al. (2013). This paper introduces Word2Vec, which is a model that can be utilized to calculate the distances between words. However, utilizing only words for comparison limits the semantic understanding of the model, as words often have different meanings based on the context of the sentence. Sentence-BERT improves on this by generating sentence embeddings, which allows for the ability to calculate the similarity between sentences (Reimers, 2019). Sentence-BERT utilizes the BERT model as developed by Devlin et al. (2018) to generate the vector embeddings.

Word2VEC and Sentence-Bert are both based on dense vector embeddings (Mikolov et al., 2013; Reimers, 2019). *Dense embeddings* contain numerous non-zero values in the vector, and as such can accurately represent interrelationships between text that are related but not exactly the same. *Sparse embeddings* in contrasts contain mostly zero values in the vector, with non-zero values only for text which is nearly identical. This makes sparse embeddings suitable for finding exact word matches. State-of-the-art models often utilize a hybrid architecture, combining both sparse and dense embeddings to give a more nuanced representation of the similarity between text. A good example of such models is BGE-m3 by Chen et al. (2024) and mGTE by Zhang et al. (2024). Both of these outperform other models when used for retrieving similar text. These models also have a sixteen times larger sequence length when compared to Sentence-BERT, allowing for the comparison of larger text (Chen et al., 2024; Reimers, 2019; Zhang et al., 2024). In addition, both BGE-m3 and mGTE have been explicitly designed to work in multilingual context (Chen et al., 2024; Zhang et al., 2024). This makes BGE-M3 and mGTE suitable for generating vector embeddings of resumes and task descriptions, as it large context length helps with the amount of text that may be in a resume, and the multilingualism helps to make it work well with resumes in different languages.

In the context of scientific research vector embeddings have previously been used to match tasks to skills, in the working paper by Amin et al. (2022). Other researchers have also focussed on improving matching candidates skills in online recruitment (Bocharova & Malakhov, 2024), as well as assisting project managers in personnel selection (Kanakaris et al., 2022), and classifying resumes (Nasser et al., 2018). This previous research indicates that vector embeddings may be a suitable technique for assessing a person's skill-fit.

2.5 Named-entity recognition

Another technique that was used in this research is named NER. NER is the task of identifying mentions of rigid designators from text belonging to predefined semantic types, such as person, location, or organization (Li et al., 2020). This is employed in this research to extract employee names from resumes and to find which languages are mentioned on the resume. This approach facilitates the creation of a structured list, such as all employee names identified in the data. While NER can be performed using human-annotated rules, this process is labour-intensive, and as a result, machine learning algorithms have been the preferred method for this task for some time (Kapetanios et al., 2013). A state-of-the-art framework for NER is Flair by Akbik et al. (2018). Various models have been built on this framework. State-of-the-art models for NER using FLAIR in English are proposed by Yu et al. (2020) and Yamada et al. (2020), both of which achieve F1 scores higher than 90. Models have also been proposed for other languages, such as the state-of-the-art Dutch model proposed by Yu et al. (2020) which achieves an F1 score higher than 90, which is of particular importance to this research as the research is being conducted in the Netherlands.

2.6 Summary

The increased focus on dynamic alignment is a result of rapidly changing markets. This is combined with a shift towards more task-based structuring, in order to respond more quickly to changing environments. Person-skill fit is a new framework that can be used for assessing the fit between the skills of persons and tasks. Various technologies may be able to help with this. The first of these is LLMs, which excel in textual understanding. A second technique that may be used is vector embeddings, which can be used to calculate the similarity between text. Moreover, NER can be used to extract structured data from unstructured data, such as resumes.

3. Systematic literature review

3.1 Stages of the SLR

In order to assess the current state of research into how technology can analyse and reallocate tasks, a systematic literature review (SLR) has been conducted. This is separate from the background information, which contained broader knowledge that would be needed to develop the prototype. Instead, the SLR looks into how task assignments are currently being restructured using LLMs in order to gain a comprehensive understanding of the current state of the research field. LLMs were specifically chosen for this as they were deemed the most promising technology for using NER to assess person-skill fit when starting the prototype development. The literature review was undertaken in distinct stages: identification and exclusion criteria, search for relevant studies, identification and selection of studies, information extraction.

3.1.1 Inclusion and exclusion criteria

Studies were eligible for inclusion if they contained both information about how task assignment can be restructured using LLMs, and contained information about the employees and the impact that work has on them. Papers that focus on the effect that LLMs have on work, which causes work to be restructured are not included, only papers that are actually about using an LLM for the process of restructuring are within the scope. Included studies could not be more than 5 years old, as technology is advancing rapidly, and it is important that the information is still relevant today. Studies had to be written in either English or Dutch and be digitally accessible.

3.1.2 Data searches and strategy

The search strategy included searching the following electronic databases: ACM Digital Library, Springerlink and IEEE Xplore. Furthermore, Google Scholar was used to find any articles that may have been missed. The entire search string that was used can be seen in Appendix A.

3.1.3 Study identification and selection

The first filtering was whether the papers were published or accepted for publication, as only published papers were included. The content type was limited to articles and conference papers. The second filtering was whether the paper was truly relevant to the topic or was erroneously included. For this, the abstract was read of all the papers to further filter whether they were relevant. The final step was to check the full articles to find whether the articles indeed turned out to be useful, and to get the useful information out of it. All papers that were not deemed relevant can be seen in Appendix B, while an overview of these steps can be seen in Appendix C. After this, forward and backward snowballing was utilized to find additional papers that might have been missed, which can be seen in Appendix D.

3.2 Findings based on SLR

As the use of LLMs to reallocate tasks is a relatively new concept, there were few peer-reviewed papers available. Below is the information that was found in the available peer-reviewed papers.

3.2.1 Impact of LLMs on Human Resource management

Previous research has indicated that LLMs can be used to make HR processes more efficient, as well as lead to better decision-making. (Budhwar et al., 2023). For instance, LLMs can shortlist CVs for hiring managers (Tinguely et al., 2023). LLMs have previously been used to create a chatbot that can provide career advice, and have also been used in practice for combining multiple separate performance reviews of employees into one comprehensive review (Al Naqbi et al., 2024; Budhwar et al., 2023). Similarly, the prototype proposed in this research utilizes LLMs to summarize the key points that might mean that someone’s skill would be a good fit for a certain task. An example of a company that already uses LLMs is DHL, which uses it to monitor the wellbeing of employees (Brown et al., 2024). Research by Wu et al. (2024) also proposes utilizing LLMs for task allocation.

LLMs can also be used for task standardization, as their understanding of language allows them to understand the intricacies of tasks (Doanh et al., 2023). This ability of LLMs to understand tasks was directly relevant to this research, as allocating tasks requires an intricate understanding of tasks. However, due to the risk of LLMs making up information, known as hallucinating, it is still important to verify the information.

3.2.2 LLM acceptance in HR

To ensure a successful implementation of LLMs on the work floor, it is crucial to consider how to make the tools acceptable and ethical (Loakimidis & Maglajlic, 2023). As the prototype uses LLMs, this is relevant to ensure acceptance by stakeholders such as HR professionals. The level of acceptance of HR-LLM systems depends on the tasks that the LLM system will be performing. The level of task difficulty is directly proportional to the level of acceptance, with routine and low cognitive complexity tasks being more easily accepted (Tinguely et al., 2023). For instance, HR decisions such as task allocation are more likely to be accepted than automatically firing someone based on an AI system's decision (Tinguely et al., 2023). Secondly, explainability has a positive impact on the acceptability of the system (Tinguely et al., 2023).

Dwivedi et al. (2023) indicate that by assuming a restructuring role, an LLM may in some ways assume a managerial role within an organization. However, as LLMs lack feelings, desires, intentions, and responsibility in the same way as humans, they are not considered team members by humans even if they are given a role part of a team. (Brown et al., 2024). This research employs various strategies to make the use of LLMs more acceptable for HR professionals, such as letting the LLM explain why someone would be a good fit for the task. Given that LLMs are seen more as tools than as team members, the prototype has been developed to be utilized as a tool, instead of feeling like interacting with a team member.

3.3 Key Insights and Implications

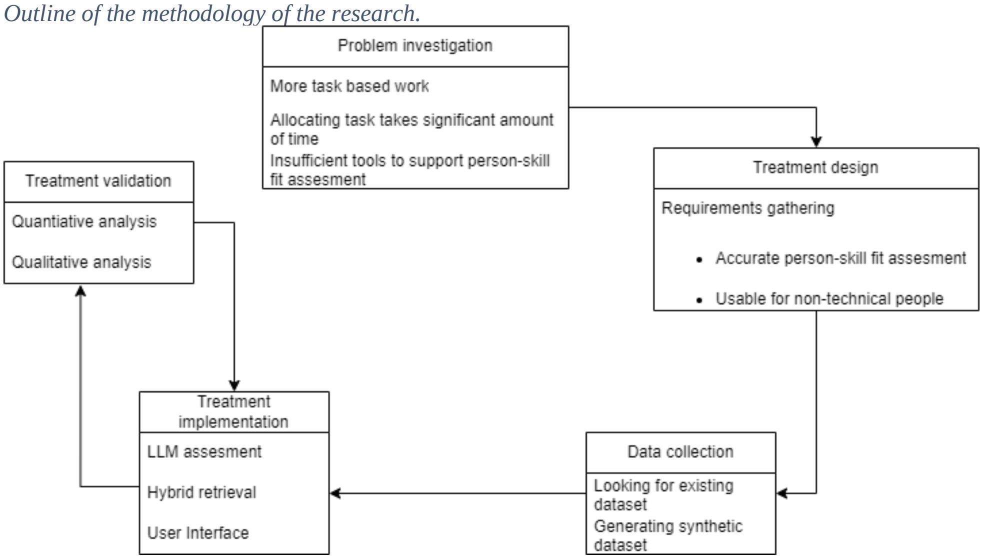
The systematic literature review has shown that LLMs have the ability to understand intricacies related to jobs and work processes, as shown by Tinguely et al. (2023), Al Naqbi et al. (2024), Doanh et al. (2023), Budhwar et al. (2023). This indicates that utilizing LLMs in supporting people within organizations in efficiently identifying the most suitable individuals for specific tasks may be a feasible approach that builds on previous research. However, utilizing LLMs in this way requires careful consideration of both ethical considerations, and considerations regarding acceptance (Budhwar et al., 2023; Ioakimidis & Maglajlic, 2023; Tinguely et al., 2023). As such, the prototype has been built to let the LLM explain its decision.

4. Research method

The research is structured according to the first five steps of the design cycle, as described by Wieringa (2014). The design cycle provides a structured approach to systematically investigate a problem, design and implement a solution, and validate its effectiveness, which made it well-suited for this study. This paragraph will give an overview of the way the research was structured, based on these cycles. Figure 1 provides an overview of the way the research was structured. It was not within the remit of this research to implement the prototype in an organizational setting during the implementation stage. Accordingly, the treatment implementation stage was not included in this research project. Consequently, the incorporation of the feedback from the stakeholders on the prototype must be addressed in future work. This also means that the implementation evaluation stage is to be addressed in future research.

Figure 1

Outline of the methodology of the research.



4.2 Problem investigation

The research started with the problem investigation. During this cycle, it was found that work is increasingly structured task-based, instead of each task being strictly linked to vocational orderings (Chalutz-Ben Gal, 2023). However, properly allocating tasks to the right people was found to take a significant amount of time (Bouajaja & Dridi, 2017). To make it easier to assess who is capable of performing a task, Chalutz-Ben Gal (2023) proposes that a platform should be developed for person–skill fit assessment, but to date, such a platform that does person-skill fit assessment based on computer analysis does not exist. In order to investigate this problem, a literature review was conducted. Based on the findings of the problem investigation, the next phase involved designing a treatment to address the challenges of supporting people within organizations in efficiently identifying the most suitable individuals for specific tasks.

4.1 Treatment design

The treatment design step of the research focussed on finding what would be needed to properly develop a platform to assist in person-skill fit assessment. This involved defining the requirements on what the prototype should be able to accomplish. Table 1 specifies the various requirements that were deemed to be important. Based on these requirements, the next phase involved gathering the necessary data to develop and test the prototype. Data collection

Table 1
A table outlining the requirement of the prototype

Requirement	Rationale
Provide accurate person-skill fit assessment	The assessment of person-skill fit should be reliable enough that people within organizations can trust it. If it is inaccurate and cannot be trusted, using the prototype may result in more instead of less work.
Be usable for non-technical people within organizations.	To make the prototype broadly usable, it cannot be assumed that each user has a lot of technical knowledge.
Give quick results	If the prototype takes too long to provide answers on person-skill fit it may not save much time for HR employees. This may also hurt the willingness to use the prototype.
Respect privacy	The prototype will be designed to handle a lot of personal information from resumes. It is both legally required and ethical to handle this information in a way that takes privacy into account.
Work offline without reliance on external services or cloud providers.	By designing the prototype in a way that can be run locally, this makes it easier for organizations to comply with privacy regulations as all data is kept in house. Furthermore, it makes the prototype more resilient as it does not rely on third parties to function properly.
Handle large amounts of resumes	Larger companies should be able to input the resumes of all their employees.
Filter resumes based on languages	If a person is able to perform a task but not speak the language that is needed, the person is still not suitable for performing the task.
Redact names on resumes	As people may be biased towards certain names, it is important that this bias can be minimized.
Assess person-skill fit multilingually	Due to globalization, companies often operate in multiple countries and employ people who speak different languages.

4.3 Data collection

The data collection of this research focussed on synthetically generating a dataset that could be used to develop and test the implementation of NLP models based on neural networks in a prototype to support people within organizations in efficiently identifying the most suitable individuals for specific tasks. As there currently is no large-scale dataset available that can be used for this purpose, the choice was made to synthetically generate a dataset of resumes.

Accordingly, a new dataset was constructed, comprising information regarding the tasks that an individual was capable of performing and the corresponding resumes. The dataset was derived from the O*NET OnLine database, version 28.3, which was released in May 2024 (National Center for O*NET Development, 2024b). The scripts utilized to generate the resumes are not contingent on this particular version, and can be employed with both newer and older iterations of the database. In order to create the requisite personal information, the Python library Faker was employed to generate non-overlapping random data. Furthermore, the LLM Mistral Nemo was used to transform the separate information into a coherent resume. Moreover, Mistral Nemo was instructed to fill in any information that might be missing to make the resume complete. The utilization of an LLM such as Mistral Nemo added a degree of randomization that meant that the resumes were not solely comprised of information from the O*NET database. This was done to make sure that the resumes would be a better representation of real-world resumes, which are not as neatly structured as the information in the O*NET database.

The O*NET database contains Standard Occupational Classification codes (SOC codes), which are unique numerical identifiers that correspond to a specific occupation or group of related occupations (National Center for O*NET Development, 2024b). Each information that was used from the O*NET database for a resume belonged to the same O*NET SOC code. When the O*NET database indicated a likelihood of a value, that value was used to weigh the chance of a certain value appearing in the resume. For instance, if 60 percent of all chief executive officers had done university, the chance that a resume of a chief executive officer would be generated stating this person completed university would be 60 percent. All variables used for the resume generation can be seen in Appendix E. The script for synthetic data generation was designed to generate a resume for each alternate job title in the O*NET database. Due to the long time it took to generate resumes, it was decided to cut off the resume generation after generating 16371 resumes, as generating resumes for all 55024 alternate titles would take more time than was available for this research (National Center for O*NET Development., 2024). Based on the dataset that was generated, the next phase involved utilizing this data to develop a prototype that could support people within organizations in efficiently identifying the most suitable individuals for specific tasks.

4.4 Treatment implementation

The treatment implementation involved generating a prototype for a platform for automatic person-skill fit assessment in order to support people within organizations in efficiently identifying the most suitable individuals for specific tasks. At first, various experiments were conducted with utilizing LLMs to judge the task-person fit. This was done by using the synthetically generated resumes and corresponding tasks from the O*NET database as a baseline. Unfortunately, it turned out that there was not a single prompt that achieved bot recall and precision over 90 percent. As such tests were done to utilize multiple prompts to assess a single resume, while this did improve the accuracy and recall, it also resulted in having to take a significant amount of time to assess the person-skill fit for each resume. This made this approach incompatible with the requirement of handling numerous resumes, and being able to provide a quick assessment. As such, utilizing vector embeddings to rank resumes based on their relevance to specific tasks was tested. This gave better results, with the most relevant resumes appearing at the top of the ranking most of the time. Furthermore, this analysis took only a few seconds for thousands of resumes.

As it was important to make the prototype easily usable for non-technical users, a web-based interface was developed that made the prototype easier to interact with. Moreover, various other functionality was added to the prototype, such as being able to filter resumes based on languages that the person is able to speak. NER was also implemented to extract the names of the people from the resumes. This also made it possible to redact the names in the resume for bias reduction. Based on the prototype that was built, the next phase involved whether this prototype helped to support people within organizations in efficiently identifying the most suitable individuals for specific tasks.

4.5 Treatment validation

The validation phase comprised both quantitative and qualitative analyses. A quantitative analysis was conducted to assess the prototype's performance, specifically examining the correlation between the assigned ranks and the degrees of task relevance. The assumption behind this was that resumes that are guaranteed to be able to perform a task should be assigned a low rank. A qualitative analysis was conducted through semi-structured interviews with the two HR professionals. During these interviews, the functionality of the prototype was demonstrated, and feedback was gathered on its utility, strengths, and areas for improvement. To assess whether the prototype is suitable for supporting people within organizations in efficiently identifying the most suitable individuals for assigning tasks to based on resumes, two interviews were conducted. The first interview was with the owner of a recruitment company specializing in the technology sector (**P1**), and the second was with the group manager of HR at an international fashion group (**P2**). P1 has extensive experience in finding people who are able to perform certain tasks that a company is looking for, while P2 has been responsible for how tasks are reallocated due to previous HR roles at various organizations. This experience in tasks allocation makes them well-suited domain experts for this research. Furthermore, the participants were selected for semi-structured interviews based on their proximity to the researcher in the Netherlands.

In addition, the interviewees were asked to evaluate synthetic resumes generated during the research study. For each resume, they indicated whether the individual was able to perform the task by assigning one of three labels: ‘yes,’ ‘maybe,’ or ‘no’. This was done to ascertain whether the ratings provided by the interviewees were consistent with those assigned by the prototype. The task that had to be assessed was: "analyze and interpret statistical data to identify significant differences in relationships among sources of information". The treatment validation also led to further steps in the treatment implementation, as the outcome of the validation revealed points for improvement both in the ranking and in the overall functionality. An example of this is summaries of the resumes, which were only implemented after the validation stage.

By following the steps of the design cycle, the research and development was structured. This chapter described the treatment design which included the problem investigation, the treatment design, data collection by generating a synthetic dataset, prototype implementation, and validation. The next chapter will give more in depth information about the development of the prototype.

5. Prototype development

As the prototype development was the most crucial aspect of the research, a distinct section is devoted to elucidating the rationale behind the design choices. The following section will describe the various steps that were taken to arrive at the final prototype.

5.1 Prototype using LLMs

As LLMs were deemed promising for automating person-skill fit assessment, a large part of the development time was spent on performing various tests using LLMs. LLMs were deemed a feasible technology due to their ability for textual understanding, as well as their ability to provide both structured output as a textual explanation of their assessment. Various strategies such as manually coming up with prompts and forcing structured output from the LLM. However, as no individual prompt was found that achieved both high recall and high precision As such, further tests were conducted using ensembles of prompts which did result in higher precision and recall. However, this still ran into two problems. As over 30 prompts had to be used per task and employee combination in order to achieve good recall and precision, this meant that each assessment took a long time. This made this method unsuitable for large datasets containing many employees, as it would take a lot of time before an answer was found. Furthermore, the ensembles were overfit and did not scale well to unseen data. It was therefore concluded that utilizing local LLMs to accurately assess such extensive data sets was not a viable option in its current form.

5.2 Prototype using vector embeddings

As the utilization of prompts with large language models did not appear to be a viable solution, an alternative approach was employed, namely to use vector embeddings. This was done by creating vector embeddings for the documents in the specified directory and saving these embeddings to a specialized vector database. Vector embeddings were created utilizing a quad embedding approach, that utilized both the dense and sparse embeddings from BGE-M3 by Chen et al. (2024), as well as dense and sparse embeddings from mGTE by Zhang et al. (2024). These embedding models were chosen for being able to handle 8 times as long embeddings as comparable models such as Sentence-BERT by Reimers (2019), allowing for inputting larger documents. Moreover, these models are both multilingual and outperform other models when used for retrieving similar text. The decision to use both BGE-M3 and mGTE was done based on the hypothesis that combining more embeddings would lead to more nuanced results. Test were conducted to confirm this hypothesis, which indicated that combining both MGE-M3 and mGTE led to resumes that were guaranteed to be able to perform a task to be ranked higher on average, than when utilizing either BGE-M3 or MGTE alone. As a vector database, Milvus by J. Wang et al. (2021) was used as it allowed for the combined usage of different embeddings, known as hybrid-retrieval, which other local open-source databases did not allow.

A benefit of utilizing embeddings is that once embeddings have been generated for the relevant documents, for example resumes, these embeddings can be reused. By only having to generate embeddings for the task description, while reusing the embeddings of the documents, this means that the assessment can be done quickly for numerous documents. Calculating the distance between vectors is a relatively quick operation compared to generating the vector embeddings, and as such the most time-consuming part of the process can be reused. Another advantage of using vector embeddings using sentence similarity models is that they are deterministic. This ensures that upon repetition of the ranking process, the same result will be obtained. This facilitates greater explicability. In contrast, alternative AI solutions, such as large language models, are not deterministic and may yield disparate results upon repetition.

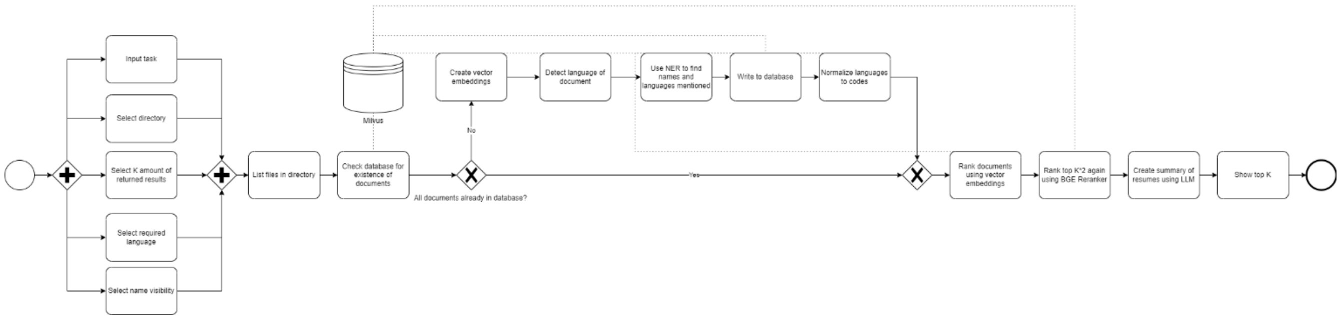
The working of the prototype is described in Figure 2. While previous versions of the prototype included more functionality, such as the ability to extract tasks from BPMN diagrams, the final prototype was designed with simplicity as a core objective, with the aim of facilitating person-task fit assessment without overwhelming the user with functions.

In order to facilitate ease of use and enhance user acceptance, the prototype was developed as a web application based on the library NiceGUI by (Schindler & Trappe, 2024). In this web-interface a user can input the task that has to be assessed, select a directory of documents that have to be assessed, specify a limit on the number of persons who have to be returned as a ranking, select required languages, and select whether the name of the persons should be visible. After the user has inputted the needed information, LlamaIndex by Liu (2022) is used to find all the documents in the specified folder and extract the text from these documents. After this, a check is done on whether documents are already in the database, as documents that are already in the database do not have to be analysed again. For documents that are not yet in the database, vector embeddings are created. After the vector embeddings were created, Flair by Akbik et al. (2018) is used for NER to extract the person names and mentioned languages from unstructured data. Specifically, English and Dutch models by Yu et al. (2020) and Yamada et al. (2020) are used for this. After this, the language that a document is in is detected using Lingua by Peter Stahl (2022). Lingua is a library that achieved state-of-the-art results for detecting the language of written text. A problem was that people may mention languages differently when referring to the same thing, for example, German and Deutsch mean the same while being different words. As such, Langcodes by Georg Krause (2014) was used to normalize these names into language codes. After this, all the new documents and extracted data such as languages and person names is added to the Milvus database.

Once all the documents have been analysed and the corresponding embeddings are in the database, the documents are ranked based on the embeddings. In order to perform this ranking, vector embeddings for the task description are generated, which are then compared to the vector embeddings of the documents. For this comparison, Reciprocal Rank Fusion is used which combines the scores of the dense and sparse embeddings from BGE-M3 by Chen et al. (2024), as well as dense and sparse embeddings from mGTE by Zhang et al. (2024). After all documents in the specified directory have been ranked, the top documents are reranked using a reranker which is more accurate than directly using the vector-embeddings, but also takes more time. Specifically, bge-reranker-v2-m3 by Chen et al. (2024) is used. Double the number of documents that the user wants to retrieve are reranked, as a document might be ranked just outside the number of documents that the user wants to retrieve while the reranker deems it important enough to place it high enough on the ranking for it to be retrieved. The choice to not rerank all documents was made to make sure that the prototype is able to rank the documents quickly, as the reranker takes significantly more time than directly using vector embeddings. After the reranker has reranked the documents, a LLM, specifically Qwen 2.5 7B by Qwen Team (2024), is used to summarize the strong points of why a person would be a good fit for the task. After the LLM is finished, the user is shown the ranked amount of specified documents, together with the languages that the persons speak. The name of the person is visible based on whether the user selected this. The prototype, as well as the relevant data, can be found at https://drive.proton.me/urls/D0ZJAMRCG8#2nxdLoBROdMz.

Figure 2

A diagram showing the working of the prototype.



In conclusion, a prototype was developed that utilizes NLP models based on neural networks to support people within organizations in assessing person-skill fit. The prototype was developed with the goal of making it easy for non-technical users to use. This prototype mainly relies on vector embeddings as using local LLMs was deemed not feasible for large datasets.

6. Evaluation

In order to evaluate the prototype, both a quantitative and qualitative analysis were conducted. The following section will present the quantitative analysis of the LLM prototype results, demonstrating why these results were insufficient for incorporating the LLM into the final prototype. Subsequently, the quantitative results will be discussed, which are intended to validate the efficacy of the ranking provided by the vector embeddings for utilization in the prototype. Thereafter, a qualitative analysis based on interviews will be presented, which was employed to verify the usability of the prototype. The section will conclude with an overview of the findings.

6.1 Quantitative Analysis

6.1.1 LLM

A series of tests were conducted to ascertain the extent to which LLMs could be utilized for the assessment of person task fit in a manner that was appropriate for large datasets. This necessitated that each individual assessment be conducted in a time-efficient manner. This leads to the first hypothesis:

First hypothesis (HA,1): A single LLM prompt can achieve both high precision (>0.8) and high recall (>0.8), resulting in an F1 score exceeding 0.8 for resume-task matching.

The efficacy of the various prompts was evaluated using a balanced dataset of 200 tasks with corresponding resumes. The highest precision was achieved by a prompt that yielded 15 true positives and no false positives, resulting in a precision of 1.0 but a recall of only 0.15 (F1 = 0.26). The prompt that achieved the highest number of true positives achieved 62 true positives and 12 false positives, resulting in a precision of 0.7 and a recall of 0.62 (F1 = 0.66). While initial attempts at combining prompts showed promise, achieving a precision of 0.95 and recall of 0.8 (F1 = 0.87) on the training data, this approach proved problematic. The combination of prompts was heavily overfit to the training data, typically retrieving only around 10 percent of true positives in unseen data. Moreover, the combination of prompts yielded unsatisfactory results, with more than 30 prompts required to achieve both high recall and precision. Given that each prompt took a few seconds to run, it was deemed infeasible to utilize prompts for datasets containing thousands of resumes. Based on these findings, the initial hypothesis must be rejected, as no single prompt could achieve the desired balance of precision and recall. Furthermore, attempts to combine prompts led to overfitting and an unacceptably long processing time. As such, it was concluded that utilizing local LLMs for person-skill fit assessment in the prototype was not a viable approach. Consequently, the remainder of the experiments employed vector embedding in lieu of LLMs.

6.2.1 Vector embeddings

6.2.1.1 Guaranteed resumes.

A statistical analysis was conducted to ascertain the extent to which the ranking could be considered an accurate indicator of skill-person fit. The dataset includes only two categories of resumes: those guaranteed to perform a task, and those for which task performance ability is unknown. Importantly, there are no instances explicitly indicating inability to perform a task. This leads to the second hypothesis:

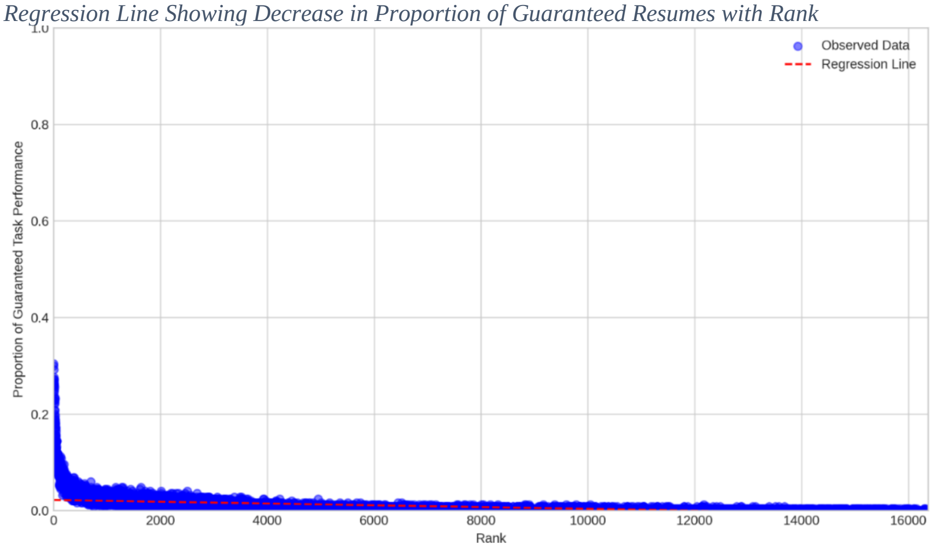
Second hypothesis ($HA,2$): The ranking aligns with task performance, such that "Guaranteed Able" resumes are ranked significantly lower (i.e., closer to rank 1) than "Unknown" resumes.

In order to analyse these hypotheses, 250 randomly sampled tasks were ranked for all 16371 resumes in the dataset. Using a Spearman Rank Correlation analysis, a strong negative correlation ($\rho = -0.7505$) was found between rank and the proportion of resumes that were guaranteed to perform the task, which was statistically significant ($p < .01$). This indicates that the ranking is found to align with task performance, with "Guaranteed Able" resumes being ranked significantly lower (i.e., closer to rank 1) than "Unknown" resumes. Using a weighted linear regression, for each unit increase in rank, there was a decrease of 1.836e-06 in the proportion of guaranteed resumes (95% CI: -1.882e-06 to -1.790e-06), which was statistically significant ($p < 0.01$). The model explains 27.0% of the variance. This further indicates that the ranking aligns with task performance, such that "Guaranteed Able" resumes being ranked significantly lower (i.e., closer to rank 1) than "Unknown" resumes.

Figure 3 presents a scatter plot of the proportion of resumes guaranteed to perform each task across various ranks, overlaid with a fitted regression line. The negative slope of the regression line clearly demonstrates that as the rank increases, the proportion of guaranteed resumes declines, which aligns with the statistical findings from the weighted linear regression model. This trend suggests that lower-ranked resumes are more likely to be guaranteed to perform a task, thereby validating the rank-ordering as a measure of task-person fit.

Figure 3

Regression Line Showing Decrease in Proportion of Guaranteed Resumes with Rank



Based on the findings from the Spearman Rank Correlation and the weighted linear regression, the second hypothesis can be accepted. Figure 4 further visualizes this relationship, with the negative slope of the regression line demonstrating the trend observed in both statistical analyses.

6.2.1.2Related resumes.

6.2.1.2.1 First degree

The dataset also includes information about job relatedness. This information allows for an analysis of whether resumes associated with jobs that are closely related to a job that is guaranteed to be able to perform the task are ranked differently than other resumes. For this analysis, resumes are categorized based on degrees of relatedness to the task (e.g., 1-degree related, 2-degree related, etc.). The objective is to test whether resumes from jobs that are a certain degree related are ranked lower (i.e., closer to rank 1) compared to all other resumes.

This was first tested for the most closely related resumes, namely those that are 1-degree related. This led to the third hypothesis:

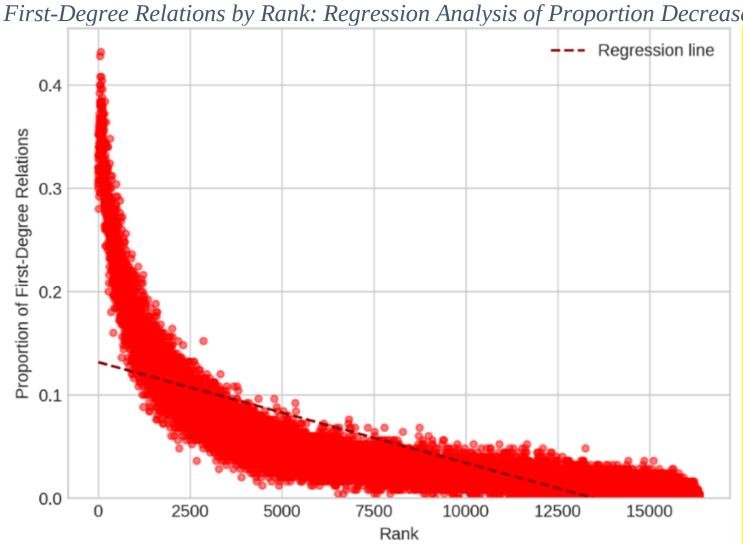
Third hypothesis ($HA,3$): The ranking is influenced by job relatedness; resumes from jobs that are 1-degree related to the task are ranked significantly lower (i.e., closer to rank 1) than all other resumes.

In order to analyze the third hypothesis, 250 randomly sampled tasks were ranked for all 16371 resumes in the dataset. Using a Spearman Rank Correlation analysis, a strong negative correlation ($\rho=-0.9119$) was identified between rank and the proportion of resumes associated with first-degree related jobs, which was statistically significant ($p<0.01$). This indicates that resumes from first-degree related jobs are ranked significantly lower (i.e., closer to rank 1) than those not associated with first-degree related jobs, demonstrating alignment between rank and job relatedness. A weighted linear regression analysis revealed a statistically significant negative relationship between rank and the proportion of first-degree related resumes ($p<0.01$). Specifically, for each unit increase in rank, there was a decrease of 9.768e-06 in the proportion of first-degree related resumes (95% CI: -9.902e-06 to -9.634e-06). The model explained 55.5% of the variance, further confirming that the ranking system prioritizes resumes from first-degree related jobs by assigning them lower ranks.

Figure 4 presents a scatter plot of the proportion of resumes associated with first-degree related jobs across various ranks, overlaid with a fitted regression line. The negative slope of the regression line demonstrates that as rank increases, the proportion of first-degree related resumes declines, aligning with the statistical findings from the weighted linear regression model. This trend indicates that lower-ranked resumes are more likely to be associated with first-degree related jobs, validating the rank-ordering as a measure of job-task relatedness.

Figure 4

First-Degree Relations by Rank: Regression Analysis of Proportion Decrease with Rank



Based on the findings from the Spearman Rank Correlation and the weighted linear regression, the third hypothesis can be accepted. These results validate the rank-ordering mechanism as an effective measure of job-task relatedness, with resumes from first-degree related jobs being prioritized appropriately by the ranking system.

6.1.2.1.2 Second degree

After having performed tests for guaranteed resumes and first-degree related, second-degree related resumes were subsequently tested. This was done to find if there was still a correlation between resumes with jobs that are more indirectly linked to a job that is guaranteed to be able to perform the task. This led to the fourth hypothesis:

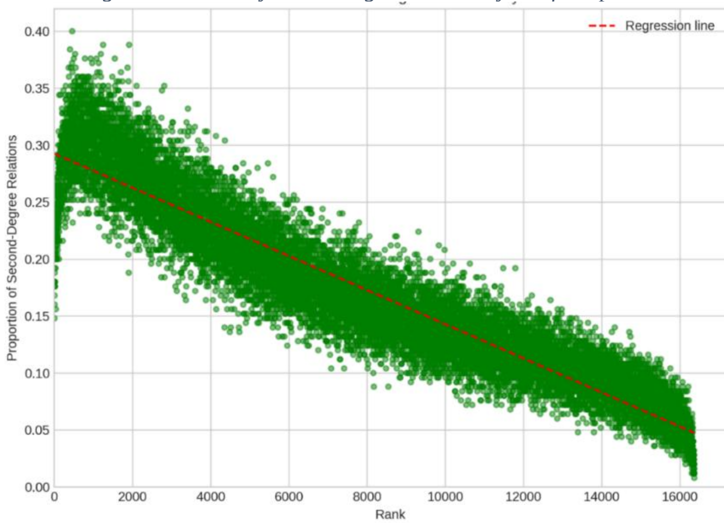
Fourth Hypothesis ($H_{A,4}$): The ranking is influenced by job relatedness; resumes from jobs that are 2-degree related to the task are ranked significantly lower (i.e., closer to rank 1) than all other resumes.

In order to analyze the fourth hyptothesis, 250 randomly sampled tasks were ranked for all 16,000 resumes in the dataset. Using a Spearman Rank Correlation analysis, a strong negative correlation ($\rho=-0.9548$) was identified between rank and the proportion of resumes associated with second-degree related jobs, which was statistically significant ($p<0.01$). This indicates that resumes from second-degree related jobs are ranked significantly lower (i.e., closer to rank 1) than those not associated with second-degree related jobs, demonstrating alignment between rank and job relatedness. A weighted linear regression analysis revealed a statistically significant negative relationship between rank and the proportion of second-degree related resumes ($p<0.01$). Specifically, for each unit increase in rank, there was a decrease of 1.497e-05 in the proportion of second-degree related resumes (95% CI: -1.505e-05 to -1.489e-05). The model explained 88.7% of the variance, further confirming that the ranking system prioritizes resumes from second-degree related jobs by assigning them lower ranks.

Figure 5 presents a scatter plot of the proportion of resumes associated with second-degree related jobs across various ranks, overlaid with a fitted regression line. The negative slope of the regression line demonstrates that as rank increases, the proportion of second-degree related resumes declines, aligning with the statistical findings from the weighted linear regression model. This trend indicates that lower-ranked resumes are more likely to be associated with second-degree related jobs, validating the rank-ordering as a measure of job-task relatedness.

Figure 5

Second-Degree Relations by Rank: Regression Analysis of Proportion Decrease with Ran



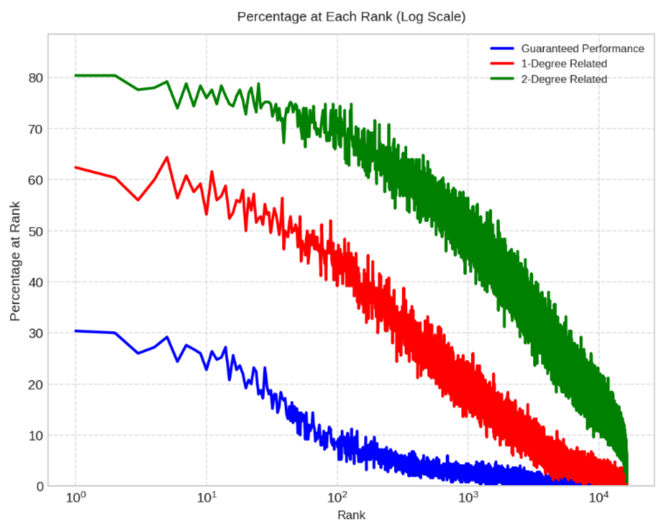
Based on the findings from the Spearman Rank Correlation and the weighted linear regression, the fourth hypothesis can be accepted. These results validate the rank-ordering mechanism as an effective measure of job-task relatedness, with resumes from second-degree related jobs being prioritized appropriately by the vector-based ranking.

6.2.1.3 Percentage at each rank

Figure 6 shows the percentage of resumes at each rank that are guaranteed to be able to perform the task or associated with related jobs that are guaranteed to be able to perform the task. Unlike previous analyses, this figure uses a cumulative approach, including closer degrees of relatedness within each category. The results indicate that lower ranks are more likely to include resumes guaranteed or related to the task, while higher ranks are less likely. A logarithmic scale is used for readability.

Figure 6

Log Scale Analysis of resumes by rank



6.2.2 Qualitative analysis

The qualitative analysis consisted of interviews with two domain experts. Both participants expressed overall satisfaction with the prototype, noting its potential as a tool for identifying suitable individuals for specific tasks. **P1** appreciated the logical ranking system, highlighting its ability to provide meaningful results even in cases where resumes lacked comprehensive detail. **P2** emphasized the prototype's relevance for large organizations, particularly for task restructuring and recruitment processes, and indicated that they would consider using the prototype in an upcoming restructuring if it were available. Both participants viewed the multilingual capabilities as a valuable feature and identified several opportunities for future improvements to enhance the prototype's utility and alignment with real-world recruitment practices. These suggested improvements are organized in Table 2.

Table 2

Summary of Suggested Improvements

Feedback/Feature	Description	Source
Location-based Filtering	Include filtering for candidates based on location to address commute distances and regional hiring practices.	P1, P2
Congruence Check	Use large language models to identify inconsistencies in resumes and apply penalties for incongruent resumes.	P1, P2
Legal Requirement Checks	Automatically retrieve and verify information from regulatory registers (e.g., BIG register for medical roles).	P1, P2
Multi-task Input and Prioritization	Allow users to input multiple tasks, assign weights, and rank candidates accordingly.	P1
Anonymous Selection	Enable anonymization of resumes to reduce bias, while ensuring optional features like photograph extraction.	P2
AI-Generated Text Identification	Flag AI-generated text or placeholders (e.g., "COMPANY NAME") to ensure resume quality.	P2
Audit Compliance	Ensure selection criteria comply with audit requirements and industry standards.	P2
Dynamic Expectations Management	Support users in adjusting expectations by reordering tasks or subtasks to reflect priority changes.	P1, P2
Cultural and Organizational Fit	Include features to assess alignment with organizational culture and responsibilities.	P2
Gender Balance in Teams	Allow filtering by gender for team balance, while adhering to legal constraints in specific regions.	P2

Table 3 compares the prototype’s rankings and normalized scores for ten synthetic resumes with expert evaluations and task relevance based on O*NET. The results indicate partial alignment between the prototype and expert judgments, particularly for highly ranked resumes and those identified as "Guaranteed" or "1st degree related" by O*NET. Notably, the owner of the technology recruitment company (P1) demonstrated a stronger alignment with both the prototype rankings and O*NET -derived relevance than the HR manager at the fashion group

(P2). This difference may stem from P1's greater familiarity with the technical nature of the task that had to be assessed. P2 exhibited more conservative judgments, reflecting less domain familiarity, and often rated resumes as "Maybe" or "No," even for resumes highly ranked by the prototype.

Table 3

*Comparison of Prototype Rankings, O*NET baseline, and Expert Ratings*

Example	1	2	3	4	5	6	7	8	9	10
Ranking (prototype)	4	5888	2	868	10	12	5047	1	11981	7137
Score (prototype)	0.7277	0.0020	0.8058	0.0242	0.6660	0.6514	0.0025	0.8378	0.0003	0.0014
Rating (O*NET)	Guaranteed	2nd degree	Guaranteed	No relation	1st degree	1st degree	2nd degree	1st degree	2nd degree	2nd degree
P1	Yes	No	Yes	Maybe	Yes	Yes	Maybe	Yes	No	No
P2	Yes	No	Maybe	Maybe	Maybe	No	No	No	No	No

6.3 Summary of Main Findings

This study aimed to develop and assess a prototype which utilizes NLP models based on neural networks for supporting people within organizations in efficiently identifying the most suitable individuals for specific tasks.

The quantitative analysis revealed that using LLMs capable of running on consumer-grade hardware is not feasible for assessing person-task fit within a short timeframe for large datasets. However, the analysis demonstrated that vector embedding models are well-suited for locally ranking large datasets of resumes. A strong negative correlation was observed between the rank assigned by the prototype and the degree of task relevance for guaranteed resumes ($\rho=-0.7505, p<0.01\rho$), first-degree related resumes ($\rho=-0.9119, p<0.01$), and second-degree related resumes ($\rho = -0.9548, p < 0.01$). Weighted linear regression models further confirmed these relationships, demonstrating statistically significant decreases in the proportions of task-relevant resumes with increasing rank. This suggests that the prototype is an effective means of prioritizing candidates who are best suited to specific tasks, thereby aligning with the objective of utilizing neural network-based NLP models to assess the person-skill fit.

The qualitative analysis, conducted through interviews with two domain experts, revealed that overall, the prototype was perceived as satisfactory. Both participants emphasized the accuracy of the prototype ranking and its potential for practical use, particularly in recruitment and task restructuring scenarios. P1 (owner of a technology recruitment company) aligned more closely with the prototype ranking and O*NET-derived relevance, reflecting greater familiarity with the technical task assessed. In contrast, P2 (HR manager in the fashion industry) demonstrated more conservative judgments, indicating less domain familiarity. Both experts also appreciated the functionality for filtering language proficiency and the potential for reducing bias through resume anonymization. Constructive feedback provided by the domain experts highlighted areas for improvement, including the addition of location-based filtering, resume congruency assessments, handling legal requirements, and support for multi-task inputs.

7. Discussion

This section discusses key limitations of the research and prototype, including methodological constraints, computational challenges, and validity threats related to synthetic data and ranking approaches. Additionally, ethical considerations, such as privacy and bias, are explored.

7.1 Limitations and threats to validity

One of the main concerns about the validity of the quantitative evaluation is selection bias, as organizations that are more receptive to change may be more likely to participate in testing the prototype. Furthermore, the findings of the quantitative analysis may have limited generalizability due to the small number of organizations involved in testing the prototype. Moreover, since the proposed task allocations have not been tested in practice, it is impossible to determine whether they are effective or simply appear to be so on paper. Additionally, the validity of the study is dependent on the accuracy of synthetic resume data. Despite care being taken to create realistic resumes, there is still a possibility that real-world resumes do not correspond to the synthetic resume. Qualitative validation was able to mitigate some of these risks, as domain experts indicated that the synthetic resumes included information that would be found in a real-world scenario. However, it should be noted that real-world resumes can vary considerably, and as such, it cannot be assumed that the generated dataset is a good representation of real-world data.

The ranking using vector embeddings produces a ranked list of resumes based on similarity scores to a task description. While this ranking offers insight into alignment, it does not definitively determine a candidate's ability to perform the task. Importantly, the meaning of a similarity score is context-dependent, as it must be evaluated relative to the scores of others in the dataset. For instance, for the task "prepare and review operational reports," a resume ranked 5000 still achieves a score of 0.01596, while for the task "pay and process claims within designated authority level," the top-ranked resume receives a score of only 0.00496, despite being guaranteed to be able to perform the task. This illustrates that scores alone are insufficient for determining suitability; instead, they must be considered in the context of overall score distributions. Another assumption is that similarity scores directly reflect task proficiency, which favours resumes focused on a single skill. While this assumption may hold in straightforward cases, it disadvantages individuals with diverse skill sets. For example, a sculptor's resume dedicated solely to sculpting may rank higher than that of a polymath like Leonardo da Vinci, despite the latter's potential superiority. For optimal resource allocation, however, individuals with a focused set of skills may be better suited for tasks directly aligned with those skills, whereas individuals with broader expertise can also be allocated to other tasks.

Moreover, the observation that guaranteed matching resumes were not always the highest-ranking resume, with other resumes often ranking higher, suggests a potential issue with the dataset's capacity to accurately reflect an individual's qualifications. The training dataset does not contain information that is ranked; however, the end product should still be a ranking, which is problematic. It would be more beneficial if a dataset could be generated that contained ranked training data from the outset, thus facilitating a more definitive confirmation of the outcomes. Besides that, minor variations in task phrasing, such as capitalization or punctuation, can also slightly affect rankings. While these effects are negligible across the entire dataset, they could influence whether specific resumes are reviewed by decision-makers, underscoring the importance of consistency in task descriptions.

A further limitation regarding the potential of implementing the prototype in organizations pertains to the availability of computational resources. The operation of NLP models based on neural networks and the management of extensive datasets necessitate substantial computational resources. For organizations lacking access to high-performance computing resources, the practical implementation of the prototype may be unfeasible. Finally, the opacity of neural networks presents a significant obstacle to transparency and explainability, which are essential for fostering trust and accountability in HR decisions. Without transparent explanations, users may be reluctant to rely on the tool, and it may be challenging to identify and rectify errors or biases in the assessments.

7.2 Ethical considerations

Although the prototype did not utilize authentic resumes, thereby significantly reducing the privacy risks associated with the prototype, the research nevertheless entails certain inherent risks. In terms of privacy, the interviewees disclosed certain company data and examples pertaining to employees who must remain anonymous. Accordingly, these examples are never directly included in the thesis in a way that would allow for the identification of the individuals or companies in question. Consequently, the companies are also not named in order to mitigate the risk of inadvertent disclosure of information that the companies would prefer to remain confidential. Although the prototype was not employed for the actual reassignment of tasks, it is designed to be capable of doing so. Should the prototype be employed for this purpose, it will have a direct impact on the careers of individuals. Consequently, any erroneous operation of the prototype will have tangible consequences in the real world. It is therefore essential that a human being remains aware of the process and that the prototype is able to provide a clear explanation of the decisions it has made. Furthermore, no bias analysis was conducted to verify that the model did not exhibit any bias in its rating. While no part of the prototype code is explicitly designed to introduce bias, it cannot be guaranteed that the used models do not inherently introduce bias. While the prototype is equipped with the functionality to render names and personally identifiable information invisible, thereby facilitating more impartial decision-making processes, this does not preclude the inclusion of names in the ranking produced by the prototype. As the vector embedding models are trained on data that includes names, such as

Wikipedia, it is probable that the model assigns certain weights to names. This may mean that names influence the assessment to a limited extent. For example, the model may associate an individual named Obama with a greater likelihood of being able to perform political functions. It would therefore be preferable if the names were excluded from the model's assessment of the resume, with a placeholder such as "PERSON NAME" substituted for the names.

8. Conclusion

This study advances the theoretical understanding of person-skill fit by operationalizing it through a prototype that utilizes NLP models based on neural networks. Practically, the prototype demonstrates potential for optimizing task allocation and recruitment processes by efficiently analysing large datasets of resumes. The following section will draw conclusions based on the research questions and list opportunities for future research.

8.1 Interpretation in Light of Research Questions

8.1.1 Sub-Research Question One: "In what ways can neural network-based NLP models be used to assess the person-skill fit?"

The research shows that vector embeddings can effectively assess person skill fit by ranking resumes by calculating similarity scores between task descriptions and resume content. The deterministic nature of the vector embeddings guarantees consistent results, which is important for transparency and reproducibility. Nonetheless, the findings highlight that similarity scores alone are insufficient for determining suitability, as they must be interpreted within the context of the dataset.

8.1.2 Sub-Research Question Two: "In what ways can neural network-based NLP models be made efficiently usable for non-technical people within organizations?"

The prototype was developed as a user-friendly web-based interface, enabling non-technical users to leverage advanced neural network-based NLP models without requiring technical expertise. Users input a task description and adjust settings, such as anonymizing names, while the system automatically processes the input and presents the ranked results in an accessible format. This streamlined design abstracts the technical complexity of the underlying models, making them intuitively usable for professionals in diverse organizational roles. Domain experts indicated satisfaction with the ease of use of the prototype. This is important as it ensures the technology is accessible to HR professionals and other stakeholders who may lack technical training.

8.1.3 Main Research Question: "How can NLP models based on neural networks support people within organizations in efficiently identifying the most suitable individuals for specific tasks?"

This research demonstrated that neural network-based NLP models, particularly those leveraging vector embeddings, provide a scalable and effective solution for ranking resumes based on task descriptions. By aligning resumes with specific tasks, the prototype developed in this research supports decision-making processes related to task allocation and recruitment. Domain experts who tested the prototype expressed satisfaction with its ranking and user-friendly design, emphasizing its potential to streamline HR processes and enhance efficiency. However, practical limitations, such as the computational demands of large-scale models and the reliance on synthetic data, must be addressed to enhance usability in real-world scenarios.

8.2 Future work

In light of the findings of this research, a number of avenues for future work have been identified with a view to enhancing the prototype and addressing the limitations observed.

Firstly, testing the prototype with real-world data from various organizations would provide more robust validation of its effectiveness and generalizability. Collaborations with companies willing to integrate the prototype into their HR processes could facilitate the generation of practical insights and contribute to the refinement of the system in accordance with the complexities inherent to real-world contexts.

Secondly, optimization of the hybrid retrieval approach is essential. An investigation into alternative vector embedding models and the fine-tuning of reranking algorithms may prove beneficial in enhancing the precision of the resume-task matching process. Moreover, integrating LLMs more efficiently, for instance by capitalizing on their capabilities in a targeted manner to conduct specific checks, could enhance the system's performance without compromising efficiency.

Thirdly, the incorporation of additional features proposed by HR professionals during the qualitative analysis would enhance the prototype's practical utility. This encompasses the implementation of location-based filtering, the handling of legal and regulatory requirements across different jurisdictions, the assessment of resume congruency, and the ability of users to input multiple tasks with assigned weights to reflect their relative importance.

Moreover, it is of paramount importance to address the ethical considerations involved. It would be beneficial for future research to focus on measuring the potential bias present in the prototype and developing suitable strategies to address it. Additionally, enhancing the explainability of the NLP models employed could facilitate user trust and enable users to identify any potential bias. A previous iteration of the prototype included the highlighting of words in the resume that corresponded to words in the task. This approach was deemed inadequate as it did not differentiate between words with the same meaning. This was an unsound approach, as not being limited to specific words was one of the key reasons why vector embeddings were used in this research. Embedding models are typically not employed for explainable AI; however, their deterministic nature facilitates the development of strategies that could enhance the explainability of these models. Further research could investigate the use of a leave-one-out sensitivity analysis, whereby various sentences are removed from the resume and the resulting change in score is observed. This approach could help identify the factors influencing the ranking and facilitate the development of a more explainable prototype.

9.1 References

Akbik, A., Blythe, D., & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. *Proceedings of the 27th International Conference on Computational Linguistics*, 1638–1649.

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review. *Sustainability*, *16*(3), 1166. https://doi.org/10.3390/su16031166

Amin, M., Bergmann, J.-P., & Campbell, Y. (2022). Using vector representations for matching tasks to skills. *HR@ RecSys*.

Audrin, B., Audrin, C., & Salamin, X. (2024). Digital skills at work–Conceptual development and empirical validation of a measurement scale. *Technological Forecasting and Social Change*, *202*, 123279.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Bocharova, M., & Malakhov, E. (2024). IMPROVING INFORMATION THEORY OF CONTEXT-AWARE PHRASE EMBEDDINGS IN HR DOMAIN. *Eastern-European Journal of Enterprise Technologies*, *131*(2).

Bouajaja, S., & Dridi, N. (2017). A survey on human resource allocation problem and its applications. *Operational Research*, *17*(2), 339–369. https://doi.org/10.1007/s12351-016-0247-8

Brown, O., Davison, R. M., Decker, S., Ellis, D. A., Faulconbridge, J., Gore, J., Greenwood, M., Islam, G., Lubinski, C., MacKenzie, N. G., Meyer, R., Muzio, D., Quattrone, P., Ravishankar, M. N., Zilber, T., Ren, S., Sarala, R. M., & Hibbert, P. (2024). Theory-Driven Perspectives on Generative Artificial Intelligence in Business and Management. *British Journal of Management*, *35*(1), 3–23. https://doi.org/10.1111/1467-8551.12788

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Lee Cooke, F., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., … Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*(3), 606–659. https://doi.org/10.1111/1748-8583.12524

Chalutz-Ben Gal, H. (2023). Person–skill fit: Why a new form of employee fit is required. *Academy of Management Perspectives*, *37*(2), 117–137.

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *ArXiv Preprint ArXiv:2402.03216*.

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*.

Doanh, D. C., Dufek, Z., Ejdys, J., Ginevičius, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., Wach, K., & Ziemba, E. (2023). Generative AI in the Manufacturing Process: Theoretical Considerations. *Engineering Management in Production and Services*, *15*(4), 76–89. https://doi.org/10.2478/emj-2023-0029

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Georg Krause. (2014, October 7). *Langcodes* . https://github.com/georgkrause/langcodes

Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology*, *48*(4), 747–773.

Hatzichronoglou, T. (1996). *Globalisation and competitiveness: relevant indicators*.

Ioakimidis, V., & Maglajlic, R. A. (2023). Neither 'Neo-Luddism' nor 'Neo-Positivism'; Rethinking Social Work's Positioning in the Context of Rapid Technological Change. *The British Journal of Social Work*, *53*(2), 693–697. https://doi.org/10.1093/bjsw/bcad081

Jiechieu, K. F. F., & Tsopze, N. (2021). Skills prediction based on multi-label resume classification using CNN with model predictions explanation. *Neural Computing and Applications*, *33*(10), 5069–5087. https://doi.org/10.1007/s00521-020-05302-x

Kanakaris, N., Giarelis, N., Siachos, I., & Karacapilidis, N. (2022). Making personnel selection smarter through word embeddings: A graph-based approach. *Machine Learning with Applications*, *7*, 100214.

Kapetanios, E., Tatar, D., & Sacarea, C. (2013). *Natural language processing: semantic aspects*. CRC Press.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*(1), 1–49.

Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 50–70.

Liu, J. (2022). *LlamaIndex*. https://doi.org/10.5281/zenodo.1234

Mandvikar, S. (2023). Augmenting intelligent document processing (IDP) workflows with contemporary large language models (LLMs). *International Journal of Computer Trends and Technology*, *71*(10), 80–91.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Mistral AI team. (2024, July 18). *Mistral NeMo*. Mistral AI. https://mistral.ai/news/mistral-nemo/

Muchinsky, P. M., & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior*, *31*(3), 268–277.

Nasser, S., Sreejith, C., & Irshad, M. (2018). Convolutional neural network with word embedding based approach for resume classification. *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, 1–6.

National Center for O*NET Development. (2024a). *Job Duties Custom List*. https://www.onetonline.org/search/task/

National Center for O*NET Development. (2024, May). *Alternate Titles*. https://www.onetcenter.org/dictionary/28.3/excel/alternate_titles.html

National Center for O*NET Development. (2024b, May). *O*NET OnLine*. https://www.onetonline.org/

Peter Stahl. (2022, January 10). *Lingua*. https://github.com/pemistahl/lingua-py/releases

Prince Thapak. (2023, July 23). *Resume-Dataset*. Hugging Face. https://huggingface.co/datasets/InferencePrince555/Resume-Dataset

Qwen Team. (2024). *Qwen2.5: A Party of Foundation Models*. https://qwenlm.github.io/blog/qwen2.5/

Reimers, N. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *ArXiv Preprint ArXiv:1908.10084*.

Rogiers, P., & Collings, D. G. (2024). The End of Jobs? Paradoxes of Job Deconstruction in Organizations. *Academy of Management Perspectives*, *ja*, amp-2022.

Schindler, F., & Trappe, R. (2024). *NiceGUI: Web-based user interfaces with Python. The nice way*. https://doi.org/10.5281/zenodo.13889918

Sheridan, T. B. (1988). Task allocation and supervisory control. In *Handbook of human-computer interaction* (pp. 159–173). Elsevier.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*(13), 1319–1350.

Tinguely, P. N., Lee, J., & He, V. F. (2023). Designing human resource management systems in the age of AI. *Journal of Organization Design*, 1–7.

Wang, C. L., & Ahmed, P. K. (2007). Dynamic capabilities: A review and research agenda. *International Journal of Management Reviews*, *9*(1), 31–51.

Wang, J., Yi, X., Guo, R., Jin, H., Xu, P., Li, S., Wang, X., Guo, X., Li, C., Xu, X., & others. (2021). Milvus: A Purpose-Built Vector Data Management System. *Proceedings of the 2021 International Conference on Management of Data*, 2614–2627.

Weller, I., Hymer, C. B., Nyberg, A. J., & Ebert, J. (2019). How matching creates value: Cogs and wheels for human capital resources research. *Academy of Management Annals*, *13*(1), 188–214.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.

Wu, W., Yang, W., Li, J., Zhao, Y., Zhu, Z., Chen, B., Qiu, S., Peng, Y., & Wang, F.-Y. (2024). Autonomous Crowdsensing: Operating and Organizing Crowdsensing for Sensing Automation. *IEEE Transactions on Intelligent Vehicles*, 1–6. https://doi.org/10.1109/TIV.2024.3355508

Yamada, I., Asai, A., Shindo, H., Takeda, H., & Matsumoto, Y. (2020). LUKE: Deep contextualized entity representations with entity-aware self-attention. *ArXiv Preprint ArXiv:2010.01057*.

Yu, J., Bohnet, B., & Poesio, M. (2020). Named entity recognition as dependency parsing. *ArXiv Preprint ArXiv:2005.07150*.

Zhang, X., Zhang, Y., Long, D., Xie, W., Dai, Z., Tang, J., Lin, H., Yang, B., Xie, P., & Huang, F. (2024). mgte: Generalized long-context text representation and reranking models for multilingual text retrieval. *ArXiv Preprint ArXiv:2407.19669*.

Appendix A: Full Search String used for SLR

The search string used was: ("task allocation*" OR "task assignment*" OR "work redesign" OR "job restructuring" OR "organizational efficiency" OR "work process redesign" OR "business process redesign") AND ("employee" OR "workforce") AND ("skill*" OR "preference*" OR "capability" OR "involvement" OR "inclusive") AND ("large language model*" OR "LLMs").The search was conducted on all text related to the article and was therefore not limited to the title. The search also required a last publication date of at last 2019.

Appendix B: Literature deemed out of scope for the SLR

title	year	journal	issn	volume	issue	pages	authors	doi
Performance management systems and multinational enterprises: Where we are and where we should go	2021	Human Resource Management		60	5	707-713	DeNisi, Angelo S.; Murphy, Kevin R.; Varma, Arup; Budhwar, Pawan	10.1002/hrm.22080
Common Good HRM: A paradigm shift in Sustainable HRM?	2020	Human Resource Management Review		30	3	100705-NA	Aust, Ina; Matthews, Brian; Muller-Camen, Michael	10.1016/j.hrmr.2019.100705
Industry 4.0 as an enabler of sustainability diffusion in supply chain: an analysis of influential strength of drivers in an emerging economy	2019	International Journal of Production Research		58	5	1505-1521	Luthra, Sunil; Kumar, Anil; Zavadskas, Edmundas Kazimieras; Mangla, Sachin Kumar; Garza-Reyes, Jo...	10.1080/00207543.2019.1660828
Human–AI collaborative decision-making as an organization design problem	2021	Journal of Organization Design		10	2	75-80	Puranam, Phanish	10.1007/s41469-021-00095-2
The Significance of Saturation	1995	Qualitative Health Research		5	2	147-149	Morse, Janice M.	10.1177/104973239500500201
New hire perceptions of their own and their employer's obligations: A study of psychological contracts	1990	Journal of Organizational Behavior		11	5	389-400	Rousseau, Denise M.	10.1002/job.4030110506
Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics	2020	Information and Organization		30	2	100301-NA	Gal, Uri; Jensen, Tina Blegind; Stein, Mari-Klara	10.1016/j.infoandorg.2020.100301
Sustainable human resource management and the triple bottom line: Multi-stakeholder strategies, concepts, and engagement	2020	Human Resource Management Review		30	3	100742-NA	Westerman, James W.; Rao, Madasu Bhaskara; Vanka, Sita; Gupta, Manish	10.1016/j.hrmr.2020.100742

Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial.	2017	JMIR mental health		4	2	0-NA	Fitzpatrick, Kathleen Kara; Darcy, Alison M.; Vierhile, Molly	10.2196/mental.7785
To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis	2022	Organization Science		33	1	126-148	Lebovitz, Sarah; Lifshitz-Assaf, Hila; Levina, Natalia	10.1287/orsc.2021.1549
FamilyBusiness.org's Editorial Guidelines for the Use of Generative AI Tools	2023	Entrepreneur and Innovation Exchange				NA-NA	Eddleston, Kimberly; Hughes, Mat; Deeds, David	10.32617/913-643eb0a448b50
Algorithms at Work: The New Contested Terrain of Control	2020	Academy of Management Annals		14	1	366-410	Kellogg, Katherine C.; Valentine, Melissa; Christin, Angèle	10.5465/annals.2018.0174
Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough?	2016	Qualitative health research		27	4	591-608	Hennink, Monique; Kaiser, Bonnie N.; Marconi, Vincent C.	10.1177/1049732316665344
Artificial Unintelligence	2018	NA				NA-NA	Broussard, Meredith	10.7551/mitpress/11022.001.0001
Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT	2023	Human Resource Management Journal		33	3	606-659	Budhwar, Pawan; Chowdhury, Soumyadeb; Wood, Geoffrey; Aguinis, Herman; Bamber, Greg J.; Beltran, ...	10.1111/1748-8583.12524
Algorithms as work designers: How algorithmic management influences the design of jobs	2022	Human Resource Management Review		32	3	100838-NA	Parent-Rocheleau, Xavier; Parker, Sharon K.	10.1016/j.hrmr.2021.100838
Enhancing the role of human resource management in corporate	2020	Human Resource		30	3	100708-NA	Stahl, Günter K.; Brewster, Chris;	10.1016/j.hrmr.2019.

sustainability and social responsibility: A multi-stakeholder, multidimensional approach to HRM		Manageme nt Review					Collings, David G.; Hajro, Aida	10070 8
Building theories from case study research.	1989	Academy of Manageme nt Review		14	4	532-550	Eisenhardt, Kathleen M.	10.546 5/amr. 1989.4 30838 5
New identities from remnants of the past: an examination of the history of beer brewing in Ontario and the recent emergence of craft breweries	2015	Business History		58	5	796-828	Lamertz, Kai; Foster, William; Coraiola, Diego M.; Kroezen, Jochem	10.108 0/0007 6791.2 015.10 65819
Performance appraisal and performance management: 100 years of progress?	2017	The Journal of applied psychology		102	3	421-433	DeNisi, Angelo S.; Murphy, Kevin R.	10.103 7/apl0 00008 5
Mediating and moderating variables of employee relations and sustainable organizations: a systematic literature review and future research agenda	2022	Internation al Journal of Organizatio nal Analysis		31	7	3023-3050	Yadav, Radha; Chaudhary, Narendra Singh; Kumar, Dharmendra; Saini, Damini	10.110 8/ijoa-12-2021-3091
When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions	2020	Organizatio nal Behavior and Human Decision Processes		160		149-167	Newman, David T.; Fast, Nathanael J.; Harmon, Derek	10.101 6/j.obh dp.202 0.03.0 08
In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms	2022	Organizatio n Science		33	1	59-82	Waardenburg, Lauren; Huysman, Marleen; Sergeeva, Anastasia	10.128 7/orsc. 2021.1 544
INSCI - Why People Use Chatbots	2017	Internet Science				377-392	Brandtzæg, Petter Bae; Følstad, Asbjørn	10.100 7/978-3-319-70284-1_30
Shattered Minds: Madmen on the Railways, 1860–80	2016	Journal of Victorian Culture		21	1	21-39	Milne-Smith, Amy	10.108 0/1355 5502.2 015.11 18851
The SAGE Handbook of Qualitative Research in Psychology	2017	NA				NA-NA	Willig, Carla; Rogers, Wendy Stainton	10.413 5/9781 52640 5555

Proposed managerial competencies for Industry 4.0 – Implications for social sustainability	2021	Technological Forecasting and Social Change		173		121080-NA	Shet, Sateesh V.; Pereira, Vijay	10.1016/j.techfore.2021.121080
The adoption of artificial intelligence in employee recruitment: The influence of contextual factors	2021	The International Journal of Human Resource Management		33	6	1125-1147	Pan, Yuan; Froese, Fabian; Liu, Ni; Hu, Yunyang; Ye, Maolin	10.1080/09585192.2021.1879206
What is it to be Critical? Teaching a Critical Approach to Management Undergraduates	2000	Management Learning		31	2	219-237	Mingers, John	10.1177/1350507600312005
Craft Imaginaries – Past, Present and Future:	2021	Organization Theory		2	1	2631787721 99114-NA	Bell, Emma; Dacin, M. Tina; Toraldo, Maria Laura	10.1177/26317877219911 41
Human resources analytics: A systematization of research topics and directions for future research	2022	Human Resource Management Review		32	2	100795-NA	Margherita, Alessandro	10.1016/j.hrmr.2020.100795
HRM institutional entrepreneurship for sustainable business organizations	2020	Human Resource Management Review		30	3	100691-NA	Ren, Shuang; Jackson, Susan E.	10.1016/j.hrmr.2019.100691
'Define, Explain, Justify, Apply' (DEJA): An analytic tool for guiding qualitative research sample size	2021	International Journal of Social Research Methodology		25	6	809-821	Mthuli, Syanda Alpheous; Ruffin, Fayth; Singh, Nikita	10.1080/13645579.2021.1941646
Guidelines for advancing theory and practice through bibliometric research	2022	Journal of Business Research		148		101-115	Mukherjee, Debmalya; Lim, Weng Marc; Kumar, Satish; Donthu, Naveen	10.1016/j.jbusres.2022.04.042
The perils of artificial intelligence in academic publishing	2022	Critical Perspectives on Accounting		87		102411-102411	Gendron, Yves; Andrew, Jane; Cooper, Christine	10.1016/j.cpa.2021.102411
FAccT - On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜	2021	Proceedings of the 2021 ACM Conference on Fairness,				610-623	Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina;	10.1145/3442188.3445922

		Accountability, and Transparency					Shmitchell, Shmargaret	
What is an adequate sample size? Operationalising data saturation for theory-based interview studies	2010	Psychology & health		25	10	1229-1245	Francis, Jill J; Johnston, Marie; Robertson, Clare; Glidewell, Liz; Entwistle, Vikki; Eccles, Mar...	10.1080/08870440903194015
Integrating strategic human capital and strategic human resource management	2017	The International Journal of Human Resource Management		29	1	34-67	Boon, Corine; Eckardt, Rory; Lepak, David P.; Boselie, Paul	10.1080/09585192.2017.1380063
Artificial Intelligence in Human Resources Management: Challenges and a Path Forward:	2019	California Management Review		61	4	15-42	Tambe, Prasanna; Cappelli, Peter; Yakubovich, Valery	10.1177/0008125619867910
We Are All Theorists of Technology Now: A Relational Perspective on Emerging Technology and Organizing	2022	Organization Science		33	1	jan-18	Bailey, Diane E.; Faraj, Samer; Hinds, Pamela J.; Leonardi, Paul M.; von Krogh, Georg	10.1287/orsc.2021.1562
How new technologies spread: Lessons from computing technologies	2013	Technology and Culture		54	2	229-261	Cortada, James W.	10.1353/tech.2013.0081
DSAA - Explaining Explanations: An Overview of Interpretability of Machine Learning	2018	2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)				80-89	Gilpin, Leilani H.; Bau, David; Yuan, Ben Z.; Bajwa, Ayesha; Specter, Michael A.; Kagal, Lalana	10.1109/dsaa.2018.00018
Green operational performance in a high-tech industry: Role of green HRM and green knowledge	2023	Journal of Business Research		160		113761-113761	Wang, Zhining; Cai, Shaohan Alan; Ren, Shuang; Singh, Sanjay Kumar	10.1016/j.jbusres.2023.113761

The Society of Algorithms	2021	Annual Review of Sociology		47	1	213-237	Burrell, Jenna; Fourcade, Marion	10.1146/annurev-soc-090820-020800
What You See is What You Get? Enhancing Methodological Transparency in Management Research	2018	Academy of Management Annals		12	1	83-110	Aguinis, Herman; Ramani, Ravi S.; Alabduljader, Nawaf	10.5465/annals.2016.0011
Sustainable careers : towards a conceptual model	2020	Journal of Vocational Behavior		117		jan-13	De Vos, Ans; van der Heijden, Beatrice; Akkermans, Jos	10.1016/j.jvb.2018.06.011
Transparency and replicability in qualitative research: The case of interviews with elite informants	2019	Strategic Management Journal		40	8	1291-1315	Aguinis, Herman; Solarino, Angelo M.	10.1002/smj.3015
Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence	2023	Academy of Management Journal		66	2	367-373	von Krogh, Georg; Roberson, Quinetta; Gruber, Marc	10.5465/amj.2023.4002
Actionable recommendations for narrowing the science-practice gap in open science	2020	Organizational Behavior and Human Decision Processes		158		27-35	Aguinis, Herman; Banks, George C.; Rogelberg, Steven G.; Cascio, Wayne F.	10.1016/j.obhdp.2020.02.007
Why Providing Humans with Interpretable Algorithms May, Counterintuitively, Lead to Lower Decision-making Performance	2022	SSRN Electronic Journal				NA-NA	DeStefano, Timothy; Kellogg, Katherine; Menietti, Michael; Vendraminelli, Luca	10.2139/ssrn.4246077
AIES - Algorithmic Hiring in Practice: Recruiter and HR Professional's Perspectives on AI Use in Hiring	2021	Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society				166-176	Li, Lan; Lassiter, Tina; Oh, Joohee; Lee, Min Kyung	10.1145/3461702.3462531

The Gender Pay Gap	2007	Academy of Management Perspectives		21	1	jul-23	Blau, Francine D.; Kahn, Lawrence M.	10.5465/amp.2007.24286161
Adoption of circular economy practices in small and medium-sized enterprises: Evidence from Europe	2022	International Journal of Production Economics		248		108496-108496	Dey, Prasanta Kumar; Malesios, Chrysovalantis; Chowdhury, Soumyadeb; Saha, Krishnendu; Budhwar, P...	10.1016/j.ijpe.2022.108496
Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: the moderating role of high-performance work systems	2021	The International Journal of Human Resource Management		33	6	1209-1236	Suseno, Yuliani; Chang, Chiachi; Hudik, Marek; Fang, Eddy S.	10.1080/09585192.2021.1931408
ChatGPT: five priorities for research.	2023	Nature		614	7947	224-226	van Dis, Eva A M; Bollen, Johan; Zuidema, Willem; van Rooij, Robert; Bockting, Claudi L	10.1038/d41586-023-00288-7
Competitive Advantage Through People	1994	California Management Review		36	2	sep-28	Pfeffer, Jeffrey	10.2307/41165742
Unlocking the value of artificial intelligence in human resource management through AI capability framework	2023	Human Resource Management Review		33	1	100899-100899	Chowdhury, Soumyadeb; Dey, Prasanta; Joel-Edgar, Sian; Bhattacharya, Sudeshna; Rodriguez-Espindol...	10.1016/j.hrmr.2022.100899
Artificial intelligence (AI)-assisted HRM: Towards an extended strategic framework	2023	Human Resource Management Review		33	1	100940-100940	Malik, Ashish; Budhwar, Pawan; Kazmi, Bahar Ali	10.1016/j.hrmr.2022.100940
EMCODIST: A Context-based Search Tool for Email Archives	2021	2021 IEEE International Conference on Big Data (Big Data)				NA-NA	Venkata, Santhilata Kuppili; Decker, Stephanie; Kirsch, David A; Nix, Adam	10.1109/bigdata52589.2021.9671832

AI-enabled recruiting: What is it and how should a manager use it?	2020	Business Horizons		63	2	215-226	Black, J. Stewart; van Esch, Patrick	10.1016/j.bushor.2019.12.001
Risky Business: How Professionals and Professional Fields (Must) Deal with Organizational Issues	2011	Organization Studies		32	10	1349-1371	Noordegraaf, Mirko	10.1177/0170840611416748
Accounting for Growth: Comparing China and India	2008	Journal of Economic Perspectives		22	1	45-66	Bosworth, Barry; Collins, Susan M.	10.1257/jep.22.1.45
The Digital Hand: How Information Technology Changed the Way Industries Worked in the United States	2006	Business History Review		80	4	755-766	Cortada, James W.	10.2307/25097268
Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize?	2021	Organization Science		32	3	856-880	Shrestha, Yash Raj; He, Vivianna Fang; Puranam, Phanish; von Krogh, Georg	10.1287/orsc.2020.1382
More Statistical and Methodological Myths and Urban Legends	2010	Organizational Research Methods		14	2	279-286	Lance, Charles E.	10.1177/1094428110391814
HRM and Performance	2004	NA				NA-NA	Paauwe, Jaap	10.1093/acprof:oso/9780199273904.001.0001
Toward a Knowledge-Based Theory of the Firm	1996	Strategic Management Journal		17		109-122	Grant, Robert M.	10.1002/smj.4250171110
Human Centric Organization Design: A Perspective from Evolutionary Psychology	2021	SSRN Electronic Journal				NA-NA	Narayanan, Jayanth; Puranam, Phanish; van Vugt, Mark	10.2139/ssrn.3959703
Advancing the sustainability agenda through strategic human resource management: Insights and suggestions for future research	2023	Human Resource Management		62	3	251-265	Ren, Shuang; Cooke, Fang Lee; Stahl, Günter K.; Fan, Di; Timming, Andrew R.	10.1002/hrm.22169
A systematic literature review on the impact of	2023	Human Resource		33	1	100857-NA	Pereira, Vijay; Hadjielias,	10.1016/j.hrm

artificial intelligence on workplace outcomes: A multi-process perspective		Management Review					Elias; Christofi, Michael; Vrontis, Demetris	r.2021.100857
Artificial Intelligence as Augmenting Automation: Implications for Employment	2021	Academy of Management Perspectives		35	4	642-659	Tschang, Feichin Ted; Almirall, Esteve	10.5465/amp.2019.0062
Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms	2022	Information Processing & Management		59	3	102940-102940	Cheng, Xusen; Zhang, Xiaoping; Cohen, Jason; Mou, Jian	10.1016/j.ipm.2022.102940
Performance Management around the Globe	2023	Performance Management Systems				360-375	DeNisi, Angelo S.; Varma, Arup; Budhwar, Pawan S.	10.4324/9781003306849-17
Is College Education Less Necessary with AI? Evidence from Firm-Level Labor Structure Changes	2022	Journal of Management Information Systems		39	3	865-905	Xue, Mei; Cao, Xing; Feng, Xu; Gu, Bin; Zhang, Yongjie	10.1080/07421222.2022.2096542
"The Main Resource is the Human"	2023	NA				NA-NA	Musser, Micah; Gelles, Rebecca; Aiken, Catherine; Lohn, Andrew	10.51593/20210071
A taxonomic foundation for evidence-based research on employee performance management	2018	European Journal of Work and Organizational Psychology		27	2	168-187	Posthuma, Richard A.; Campion, Michael C.; Campion, Michael A.	10.1080/1359432x.2018.1438411
Human-Algorithm Ensembles	2021	SSRN Electronic Journal				NA-NA	Choudhary, Vivek; Marchetti, Arianna; Shrestha, Yash Raj; Puranam, Phanish	10.2139/ssrn.3902402
Leader-member exchange in the age of remote work	2022	Human Resource Development International		25	2	219-230	Varma, Arup; Jaiswal, Akanksha; Pereira, Vijay; Kumar, Y L N	10.1080/13678868.2022.2047873
The contribution of new technology to economic	2010	Revista de Historia		28	3	409-440	Crafts, Nicholas	10.1017/s021

growth: lessons from economic history		Económica / Journal of Iberian and Latin American Economic History						2610910000157
Artificial intelligence-driven supply chain resilience in Vietnamese manufacturing small- and medium-sized enterprises	2023	International Journal of Production Research				jan-40	Dey, Prasanta Kumar; Chowdhury, Soumyadeb; Abadie, Amelie; Vann Yaroson, Emilia; Sarkar, Sobhan	10.1080/00207543.2023.2179859
The Organization of Craft Work	2018	NA				NA-NA	Bell, Emma; Mangia, Gianluigi; Taylor, Scott; Toraldo, Maria Laura	10.4324/9781315205861
The productivity paradox of information technology	1993	Communications of the ACM		36	12	66-77	Brynjolfsson, Erik	10.1145/163298.163309
Artificial Intelligence and Management: The Automation–Augmentation Paradox	2021	Academy of Management Review		46	1	192-210	Raisch, Sebastian; Krakowski, Sebastian	10.5465/amr.2018.0072
May the bots be with you! Delivering HR cost-effectiveness and individualised employee experiences in an MNE	2020	The International Journal of Human Resource Management		33	6	1148-1178	Malik, Ashish; Budhwar, Pawan; Patel, Charmi; Srikanth, N. R.	10.1080/09585192.2020.1859582
Towards a standard for identifying and managing bias in artificial intelligence	2022	NA				NA-NA	Schwartz, Reva; Vassilev, Apostol; Greene, Kristen; Perine, Lori; Burt, Andrew; Hall, Patrick	10.6028/nist.sp.1270
The Skill Content of Recent Technological Change: An Empirical Exploration	2003	The Quarterly Journal of Economics		118	4	1279-1333	Autor, D. H.; Levy, F.; Murnane, R. J.	10.1162/003355303322552801
How will Language Modelers like ChatGPT Affect Occupations and Industries?	2023	SSRN Electronic Journal				NA-NA	Felten, Edward W.; Raj, Manav;	10.2139/ssrn.4375268

							Seamans, Robert	
When Digital Technologies Enable and Threaten Occupational Identity: The Delicate Balancing Act of Data Scientists	2021	MIS Quarterly		45	3	1087-1112	Vaast, Emmanuelle; Pinsonneault, Alain	10.25300/misq/2021/16024
Algorithms of Oppression	2020	NA				NA-NA	Noble, Safiya Umoja	10.18574/nyu/9781479833641.001.0001
Awe-inspiring advancements in AI: The impact of ChatGPT on the field of Organizational Behavior	2023	Journal of Organizational Behavior		44	2	177-179	Dasborough, Marie T.	10.1002/job.2695
Craft, magic and the re-enchantment of the world	2017	European Management Journal		35	3	285-296	Suddaby, Roy; Ganzin, Max; Minkus, Alison	10.1016/j.emj.2017.03.009
The OSQE Model: The AI Cycle Against the Shortage of Skilled professionals:A Holistic Solution Approach Based on Artificial Intelligence in Times of Demographic Change	2023	NA				NA-NA	Tcharnetsky, Marina; Vogt, Florian	10.20944/preprints202302.0069.v1
Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover	2022	The International Journal of Human Resource Management		34	14	2732-2764	Chowdhury, Soumyadeb; Joel-Edgar, Sian; Dey, Prasanta Kumar; Bhattacharya, Sudeshna; Kharlamov, A...	10.1080/09585192.2022.2066981
Artificial Unintelligence: How Computers Misunderstand the World	2022	The European Legacy		27	7	860-862	Schweizer, Karl W.	10.1080/10848770.2022.2110366
The Great Divergence	2009	NA				NA-NA	Pomeranz, Kenneth	10.2307/j.ctt7sv80
Perceived usefulness, perceived ease of use, and user acceptance of information technology	1989	MIS Quarterly		13	3	319-340	Davis, Fred D.	10.2307/249008
Artificial intelligence – challenges and	2022	The Internation		33	6	1065-1097	Budhwar, Pawan; Malik,	10.1080/0958

opportunities for international HRM: a review and research agenda		al Journal of Human Resource Manageme nt					Ashish; De Silva, M. T. Thedushika; Thevisuthan, Praveena	5192.2 022.20 35161
Making Sense of (Mis)Matched Frames of Reference: A Dynamic Cognitive Theory of (In)Stability in HR Practices	2021	Industrial Relations: A Journal of Economy and Society		61	3	268-289	Budd, John W.; Pohler, Dionne; Huang, Wei	10.111 1/irel.1 2275
Humanizing work in the digital age: Lessons from socio-technical systems and quality of working life initiatives	2022	Human Relations		75	8	1461-1482	Guest, David; Knox, Angela; Warhurst, Chris	10.117 7/0018 72672 21092 674
Firm Resources and Sustained Competitive Advantage	1991	Journal of Manageme nt		17	1	99-120	Barney, Jay	10.117 7/0149 20639 10170 0108
How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability	2006	Field Methods		18	1	59-82	Guest, Greg; Bunce, Arwen; Johnson, Laura	10.117 7/1525 822x0 52799 03
Can HR adapt to the paradoxes of artificial intelligence?	2022	Human Resource Manageme nt Journal		32	4	729-742	Charlwood, Andy; Guenole, Nigel	10.111 1/1748 - 8583.1 2433
Bias and Productivity in Humans and Machines	2019	SSRN Electronic Journal				NA-NA	Cowgill, Bo	10.213 9/ssrn. 34337 37
Artificial intelligence and people management: A critical assessment through the ethical lens	2023	Human Resource Manageme nt Review		33	1	10092 3- 10092 3	Varma, Arup; Dawkins, Cedric; Chaudhuri, Kaushik	10.101 6/j.hrm r.2022. 10092 3
Elevating Employment Practices in Agricultural Corporations with Large Language Models and AI	2023	2023 Tenth Internation al Conference on Social Networks Analysis, Manageme nt and Security (SNAMS)	9,80 E+1 2			1-mei	Abu-Shanab, Samia A and Mughaid, Ala and AlZu'bi, Shadi	NA
Crowd Sensing Intelligence for ITS: Participants, Methods, and Stages	2023	IEEE Transaction s on	2379 - 8904				Zhao, Yong and Hu, Cong and Zhu, Zhengqiu and	NA

		Intelligent Vehicles					Qiu, Sihang and Chen, Bin and Jiao, Peng and Wang, Fei-Yue	
Designing for Hybrid Intelligence: A Taxonomy and Survey of Crowd-Machine Interaction	2023	Applied Sciences	2076-3417	13	4	2198	Correia, António and Grover, Andrea and Schneider, Daniel and Pimentel, Ana Paula and Chaves, Ramon and De Almeida, Marcos Antonio and Fonseca, Benjamim	NA
Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors	2023	The Twelfth International Conference on Learning Representations					Chen, Weize and Su, Yusheng and Zuo, Jingwei and Yang, Cheng and Yuan, Chenfei and Chan, Chi-Min and Yu, Heyang and Lu, Yaxi and Hung, Yi-Hsin and Qian, Chen	NA
Aircraft Line Maintenance Scheduling Using Simulation and Reinforcement Learning	2023	2023 Winter Simulation Conference (WSC)	9,80E+12			600-611	Widmer, Simon and Shaukat, Syed and Wu, Cheng-Lung	NA
Diversity Climate Perceptions							Farooqi, Rahela	NA
Enhancing conversational agents for successful operation: A multi-perspective evaluation approach for continuous improvement	2023	Electronic Markets	1019-6781	33	1	39	Lewandowski, Tom and Kučević, Emir and Leible, Stephan and Poser, Mathis and Böhmann, Tilo	NA
Kt-bt: A framework for knowledge transfer through behavior trees in multirobot systems	2023	IEEE Transactions on Robotics	1552-3098				Venkata, Sanjay Sarma Oruganti and Parasuraman, Ramviyas and Pidaparti, Ramana	NA

Tool, Teammate, Superintelligence: Identification of ChatGPT-Enabled Collaboration Patterns and their Benefits and Risks in Mutual Learning	2024		99813317 5				Cheng, Xusen and Zhang, Shuang	NA
The Impact of AIGC on Organizational Knowledge Creation: From the Perspective of Adaptive Structuration Theory	2023	2023 IEEE International Conference on Data Mining Workshops (ICDMW)	9,80 E+1 2			1469-1476	Zhang, Xi and Liu, Ziyue and Cheng, Yihang and Wang, Xuyan and Wang, Zhe	NA
Exploring Challenges and Opportunities of Wearable Robots: A Comprehensive Review of Design, Human-Robot Interaction and Control Strategy	2023	APSIPA Transactions on Signal and Information Processing	2048-7703	12	1		Chen, Lufeng and Xie, Hongqin and Liu, Zicheng and Li, Bin and Cheng, Hong	NA
E-Government 3.0: An AI Model to Use for Enhanced Local Democracies	2023	Sustainability	2071-1050	15	12	9572	Vrabie, Catalin	NA
A cognitive method for comparing and elaborating on technology frames	2023	British Journal of Management	1045-3172				Ghobadi, Shahla and Mathiassen, Lars	NA
How empty is Trustworthy AI? A discourse analysis of the Ethics Guidelines of Trustworthy AI	2024	Critical Policy Studies	1946-0171			jan-18	Stamboliev, Eugenia and Christiaens, Tim	NA
Engaging student opinions on vaccine development innovation: Experiences from a “Shark Tank” project	2024	Clinical and Translational Science	1752-8054	17	2	e13723	Barrett, Jeffrey S and Skolnik, Jeffrey M and Ingram, Mary and Kuo, Yin-Ming and Metzloff, Ann E and Jin, Ruizhe and Wu, Yuanhan and Kroushl, Nik	NA
Chatbot Persona Selection Methodology for Emotional Support	2023	2023 62nd Annual Conference of the Society of Instrument and Control	4907764804			333-338	Yorita, Akihiro and Egerton, Simon and Chan, Carina and Kubota, Naoyuki	NA

		Engineers (SICE)						
Open Radio Access Networks for Smart IoT Systems: State of Art and Future Directions	2023	Future Internet	1999-5903	15	12	380	Musa, Abubakar Ahmad and Hussaini, Adamu and Qian, Cheng and Guo, Yifan and Yu, Wei	NA
Digital & cognitive corporate reality	2023	Infocommunications Journal	2061-2079	15		2-okt	Kő, Andrea and Szabó, Ildikó and Csapó, Ádám Balázs and Kovács, Tibor and Lőrincz, László	NA
Augmenting intelligent document processing (IDP) workflows with contemporary large language models (LLMs)	2023	International Journal of Computer Trends and Technology		71	10	80-91	Mandvikar, Shreekant	NA
Analysis And Evaluation Of ChatGPT-Induced HCI Shifts In The Digitalised Translation Process	2023	2023) Proceedings of Hit-IT: The International Conference on Human-Informed Translation and Interpreting Technology. Naples: Incoma				227-67	Sánchez-Gijón, Pilar and Palenzuela-Badiola, Leire	NA
Employing ChatGPT for the Management of Businesses and Decision Making in the Era of AI	2023	Advances in Business Information Systems and Analytics				15-27	Jain, Nikita; Dhingra, Rekha; Bhardwaj, Deepa	10.4018/979-8-3693-0815-8.ch002
Artificial intelligence (AI) futures: India-UK collaborations emerging from the 4th Royal Society Yusuf Hamied workshop	2023	International Journal of Information Management				102725-102725	Dwivedi, Yogesh K.; Hughes, Laurie; Bhadeshia, Harshad K.D.H.; Ananiadou,	10.1016/j.ijinfomgt.2023.102725

							Sophia; Cohn, Anthony G...	
Benchmarking Generative AI: A Comparative Evaluation and Practical Guidelines for Responsible Integration into Academic Research	2023	SSRN Electronic Journal				NA-NA	Morande, Swapnil	10.2139/ssrn.4571867
Potential Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Architectural Engineering	2023	SSRN Electronic Journal				NA-NA	Rane, Nitin	10.2139/ssrn.4607767
Exploring the Impact of Chatgpt on Art Creation and Collaboration: Benefits, Challenges and Ethical Implications	2024	NA				NA-NA	Zhu, Sijin; Wang, Zheng; Zhuang, Yuan; Jiang, Yuyang; Guo, Mengyao; Zhang, Xiaolin; Gao, Ze	10.2139/ssrn.4681228
A Bibliometric Analysis of Artificial Intelligence and Human Resource Management Studies	2023	Exploring the Intersection of AI and Human Resources Management				85-117	Morshidi, Azizan Bin; Satar, Nurhizam Safie Mohd; Azizan, Azueryn Annatassia Dania Aqeela; Idris,...	10.4018/979-8-3693-0039-8.ch006
Reactions of applicants with disabilities to technology-enabled recruitment and selection: A research agenda	2023	International Journal of Selection and Assessment				NA-NA	Fisher, Sandra L.; Connelly, Catherine E.; Bonaccio, Silvia	10.1111/ijsa.12456
Adaptation of Artificial Intelligence Literacy Scale into Turkish	2023	Bilgi ve İletişim Teknolojileri Dergisi		5	2	172-190	KARAOĞLAN YILMAZ, Fatma Gizem; YILMAZ, Ramazan	10.53694/bited.1376831
Small and medium-sized enterprises as technology innovation intermediaries in sustainable business ecosystem: interplay between AI adoption, low carbon management and resilience	2023	Annals of Operations Research				NA-NA	Roux, Mélanie; Chowdhury, Soumyadeb; Kumar Dey, Prasanta; Vann Yaroson, Emilia; Pereira, Vijay; A...	10.1007/s10479-023-05760-1
SELECCIÓN BIBLIOGRÁFICA	2023	RVGP 25			25	114-121	,	10.47623/ivap-

								rvgp.25.2023.08
Revisiting the role of HR in the age of AI: bringing humans and machines closer together in the workplace	2024	Frontiers in Artificial Intelligence		6		NA-NA	Fenwick, Ali; Molnar, Gabor; Frangos, Piper	10.3389/frai.2023.1272823
Analyzing the PESTEL Factors on ChatGPT and Its Impact on the Population	2023	Artificial Intelligence Applications Using ChatGPT in Education				126-137	,	10.4018/978-1-6684-9300-7.ch012
A systematic literature review exploring and linking circular economy and sustainable development goals in the past three decades (1991–2022)	2023	International Journal of Production Research		62	4	1399-1433	Vann Yaroson, Emilia; Chowdhury, Soumyadeb; Mangla, Sachin Kumar; Dey, Prasanta; Chan, Felix T. S...	10.1080/00207543.2023.2270586
Metaverse-infused academic libraries: a glimpse into the future	2023	Library Hi Tech News		40	10	17-19	Amzat, Omolara Basirat; Adewojo, Akinade Adebowale	10.1108/lhtn-10-2023-0187
Contribution of ChatGPT and Similar Generative Artificial Intelligence for Enhanced Climate Change Mitigation Strategies	2024	SSRN Electronic Journal				NA-NA	Rane, Nitin; Choudhary, Saurabh; Rane, Jayesh	10.2139/ssrn.4681720
Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review	2024	Sustainability		16	3	1166-NA	Al Naqbi, Humaid; Bahroun, Zied; Ahmed, Vian	10.3390/su16031166
Multipath Serially-Mediating Mechanisms of Employee Attitudes and Behaviors	2023	Advances in Business Information Systems and Analytics				jan-22	Katou, Anastasia A.; Dhiman, Mohinder Chand; Vayona, Anastasia; Gianni, Maria	10.4018/979-8-3693-1902-4.ch001
Can a Computer Outfake a Human?	2023	NA				NA-NA	Phillips, Judith Jane; Robie, Chet	10.2139/ssrn.4533675

From Data Scarcity to Data Abundance: Crafting Synthetic Survey Data in Management Accounting using ChatGPT	2023	SSRN Electronic Journal				NA-NA	Motoki, Fabio; Monteiro, Januário; Malagueño, Ricardo; Rodrigues, Victor	10.2139/ssrn.4595896
Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations	2023	BenchCouncil Transactions on Benchmarks, Standards and Evaluations		3	3	100140-100140	Raj, Rohit; Singh, Arpit; Kumar, Vimal; Verma, Pratima	10.1016/j.tbench.2023.100140
Theory-Driven Perspectives on Generative Artificial Intelligence in Business and Management	2024	British Journal of Management		35	1	mrt-23	Brown, Olivia; Davison, Robert M.; Decker, Stephanie; Ellis, David A.; Faulconbridge, James; Gore...	10.1111/1467-8551.12788
ChatGPT in finance: Applications, challenges, and solutions	2024	Heliyon		10	2	e24890-e24890	Khan, Muhammad Salar; Umer, Hamza	10.1016/j.heliyon.2024.e24890
Generative AI in the Manufacturing Process: Theoretical Considerations	2023	Engineering Management in Production and Services		15	4	76-89	Doanh, Doung Cong; Dufek, Zdenek; Ejdys, Joanna; Ginevičius, Romualdas; Korzynski, Pawel; Mazurek...	10.2478/emj-2023-0029
Are both generative AI and ChatGPT game changers for 21st-Century operations and supply chain excellence?	2023	International Journal of Production Economics		265		109015-109015	Fosso Wamba, Samuel; Queiroz, Maciel M.; Chiappetta Jabbour, Charbel Jose; Shi, Chunming (Victor)	10.1016/j.ijpe.2023.109015
Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations	2023	Cogent Education		11	1	NA-NA	Darwin, NA; Rusdin, Diyenti; Mukminatien, Nur; Suryati, Nunung; Laksmi,	10.1080/2331186x.2023.2290342

							Ekaning D.; Marzuki, NA	
Public attitudes and sentiments toward ChatGPT in China: A text mining analysis based on social media	2024	Technology in Society		76		102442-102442	Lian, Ying; Tang, Huiting; Xiang, Mengting; Dong, Xuefan	10.1016/j.techsoc.2023.102442
Designing human resource management systems in the age of AI	2023	Journal of Organization Design		12	4	263-269	Tinguely, Patrick Nicolas; Lee, Junghyun; He, Vivianna Fang	10.1007/s41469-023-00153-x
Ethics and international business research: Considerations and best practices	2024	International Business Review		33	1	102207-102207	Miller, Stewart R.; Moore, Fiona; Eden, Lorraine	10.1016/j.ibusrev.2023.102207
Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Human Resource Management	2023	SSRN Electronic Journal				NA-NA	Rane, Nitin	10.2139/ssrn.4603230
Development of the potential of the digital economy of Russian regions through artificial intelligence humanisation	2023	Humanities and Social Sciences Communications		10	1	NA-NA	Ekimova, Ksenia V.	10.1057/s41599-023-02444-w
ChatGPT and Similar Generative Artificial Intelligence (AI) for Building and Construction Industry: Contribution, Opportunities and Challenges of Large Language Models for Industry 4.0, Industry 5.0, and Society 5.0	2023	SSRN Electronic Journal				NA-NA	Rane, Nitin	10.2139/ssrn.4603221
A multidimensional approach towards addressing existing and emerging challenges in the use of ChatGPT	2023	AI and Ethics				NA-NA	Khan, Muhammad Salar	10.1007/s43681-023-00360-y
Navigating the Future of Digital Transformation and Leadership	2024	Employee Uncertainty Over Digital Transformation				189-208	Matsunaga, Masaki	10.1007/978-981-99-8409-1_5
Impact of carbon offset perceptions on greenwashing: Revealing intentions and strategies	2024	Industrial Marketing Management		117		304-320	Abadie, Amelie; Chowdhury, Soumyadeb;	10.1016/j.indmarman.2024

through an experimental approach							Mangla, Sachin Kumar; Malik, Shaily	.01.001
Real Research with Fake Data: A Tutorial on Conducting Computer Simulation for Research and Teaching	2023	Organizational Research Methods				NA-NA	Sturman, Michael C.	10.1177/10944281231215024
Intelligent Manufacturing through Generative Artificial Intelligence, Such as ChatGPT or Bard	2024	SSRN Electronic Journal				NA-NA	Rane, Nitin; Choudhary, Saurabh; Rane, Jayesh	10.2139/ssrn.4681747
Generative artificial intelligence and academia: Implication for research, teaching and service	2023	Management Learning		54	5	597-604	Barros, Amon; Prasad, Ajnesh; Śliwa, Martyna	10.1177/13505076231201445
Assessing the Net Benefits of Generative Artificial Intelligence Systems for Wealth Management Service Innovation: A Validation of the Delone and Mclean Model of Information System Success	2023	Transfer, Diffusion and Adoption of Next-Generation Digital Technologies				56-67	Kulkarni, Mugdha Shailendra; Pramod, Dhanya; Patil, Kanchan Pranay	10.1007/978-3-031-50192-0_6
Transforming Human Resources With AI		Industrial Applications of Big Data, AI, and Blockchain				254-299	Massoud, Mazen Fawaz; Maaliky, Bassel; Fawal, Abir; Mawllawi, Allam; Yahkni, Fadlallah	10.4018/979-8-3693-1046-5.ch011
The role of generative artificial intelligence (GAI) in customer personalisation (CP) development in SMEs: a theoretical framework and research propositions	2023	Industrial Artificial Intelligence		1	1	NA-NA	Abrokwah-Larbi, Kwabena	10.1007/s44244-023-00012-4

Appendix C: Study Filtering and Selection Process in SLR

Table 4

Found papers and their sources

Source	ACM Digital Library	IEEE Xplore	SpringerLink	Google scholar
Total hits	49	0	1 (also found by Google Scholar)	90
After filtering date and type of paper	1	0	1 (also found by Google Scholar)	24
Still relevant after filtering abstract	0	0	1 (also found by Google Scholar)	13
Relevant after reading full article.	0	0	1 (also found by Google Scholar)	3

Appendix D Forward and backward snowballing in SLR

Table 5

Results of forward and backward snowballing

Source	Backward citation chasing	Forward citation chasing
Total hits	106	44
After filtering date and type of paper	67	25
Still relevant after filtering abstract	22	7
Relevant after reading the full article.	2	3

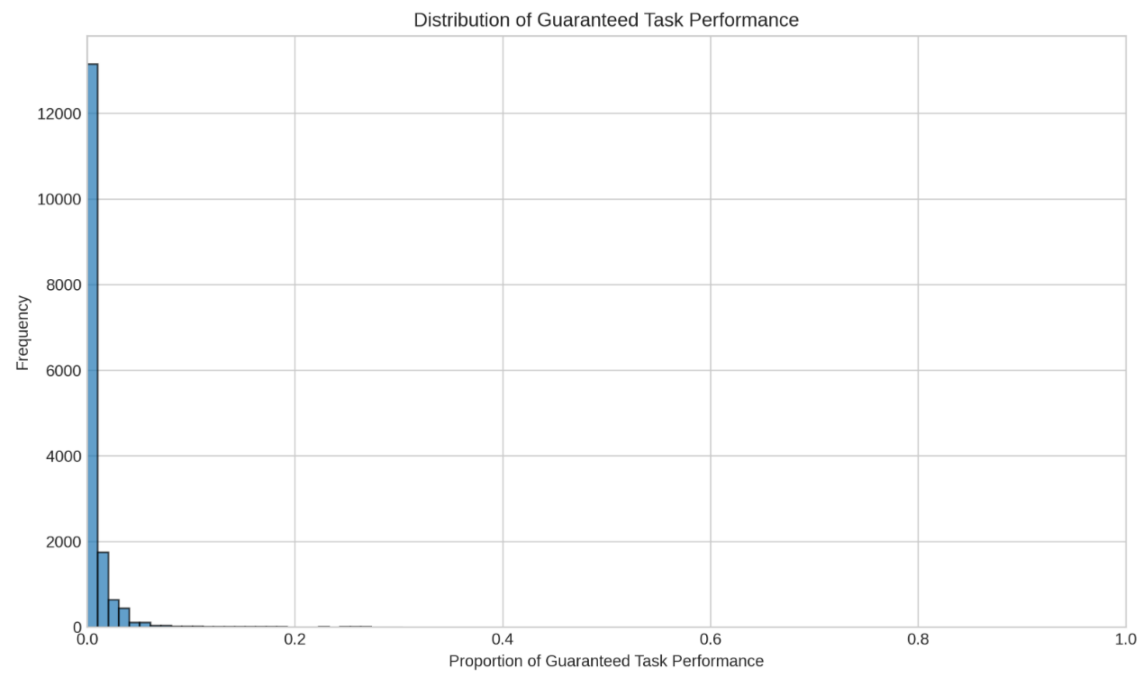
Appendix E: Resume generation details

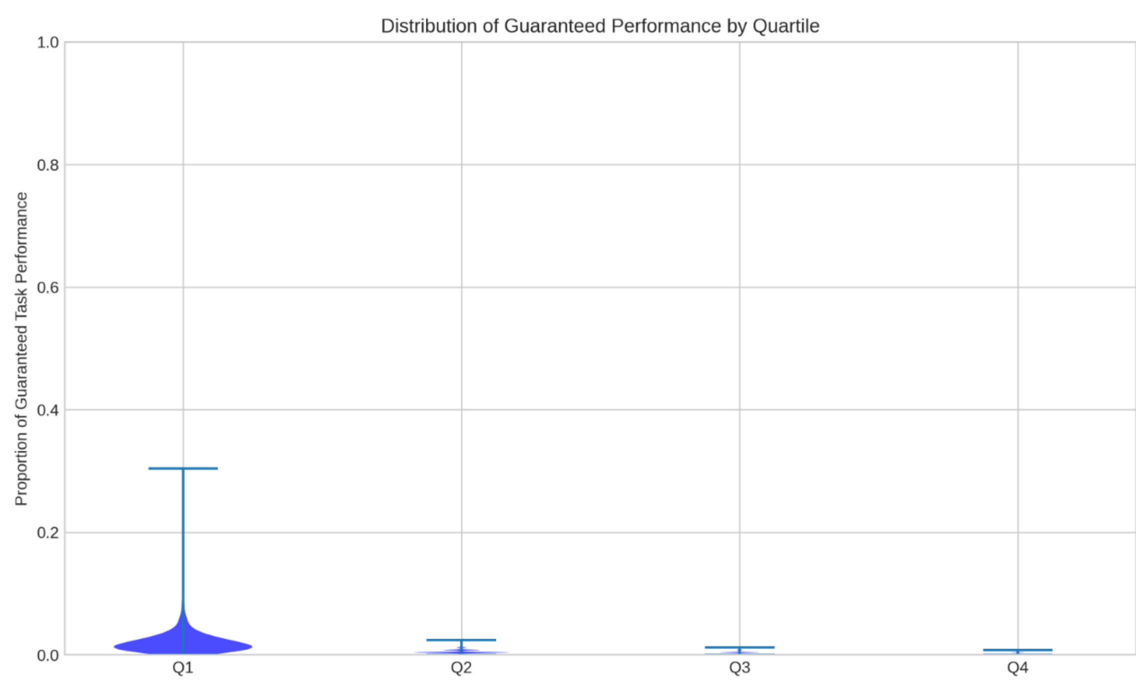
Table 6

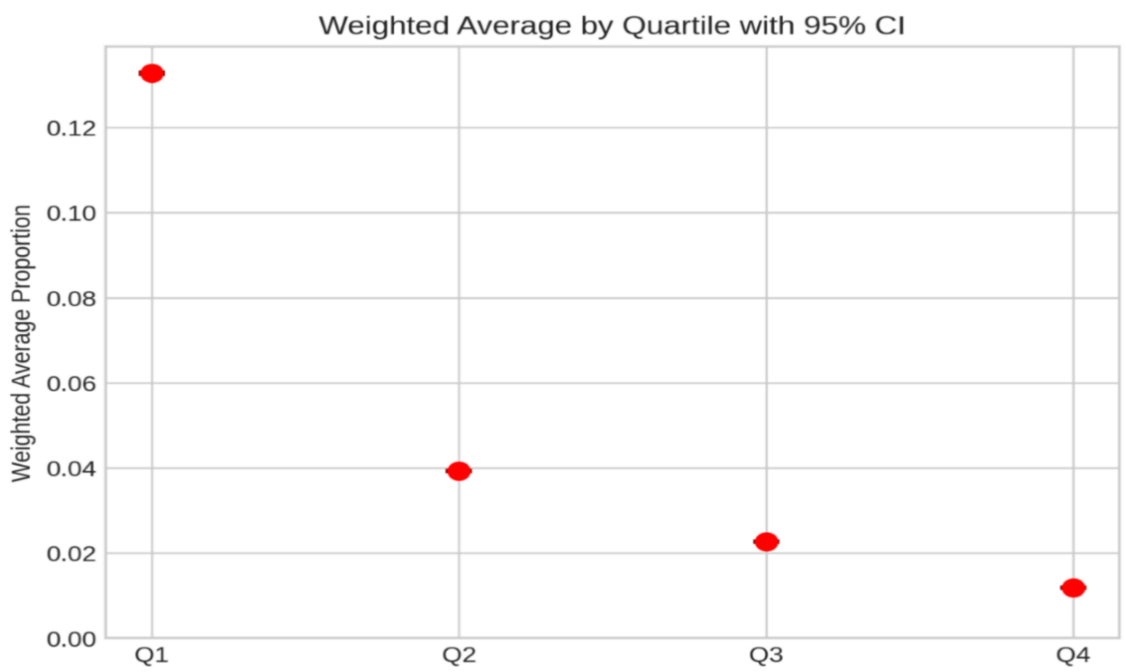
A description of how the resumes are generated

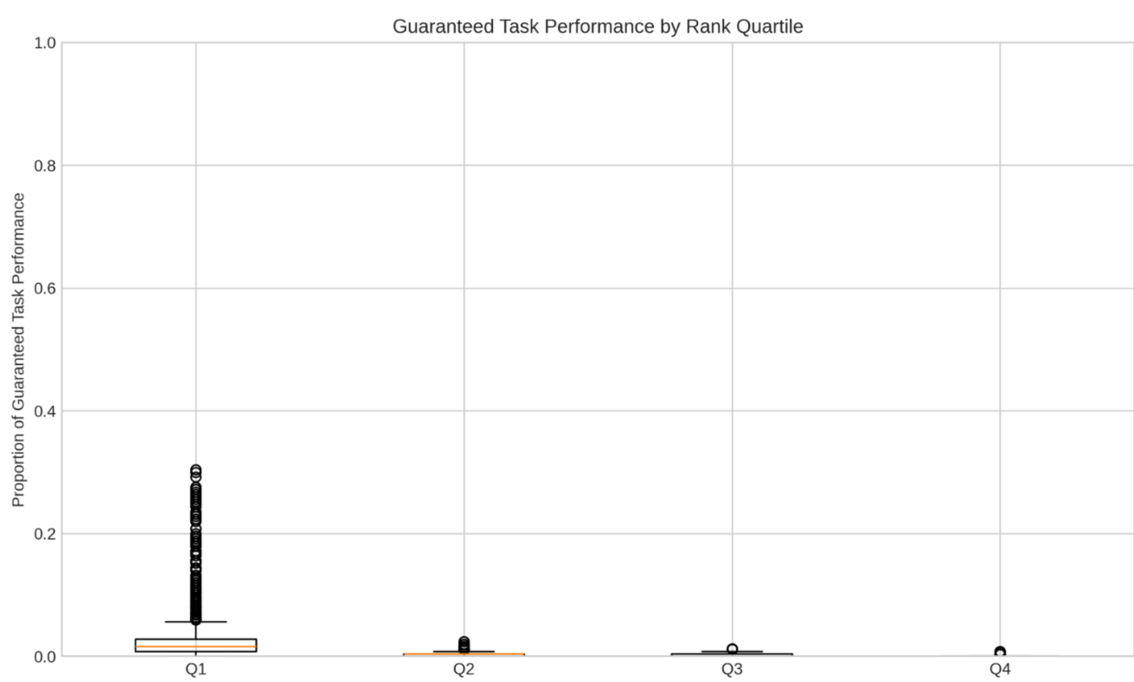
Resume Part	Source	Selection Details
Name	Faker	Exactly one value, generated randomly
Address	Faker	Exactly one value, generated randomly
Telephone number	Faker	Exactly one value, generated randomly
Email address	Faker	Exactly one value, generated randomly
Job title	Title from O*NET database	Exactly one value, chosen based on SOC code
Job description	O*NET database	Exactly one value, chosen based on SOC code
Detailed work activities	O*NET database	Up to six values, selected randomly from a list
Second-to-last position	Alternate titles from O*NET	Exactly one value, selected randomly if multiple options exist
Skills	O*NET database	Up to six values, selected randomly from a weighted list
Knowledge areas	O*NET database	Up to six values, selected randomly from a weighted list
Related technologies	O*NET database	Up to six values, selected randomly from a list
Tools	O*NET database	Up to six values, selected randomly from a list
Educational details	O*NET database	Exactly one value, selected randomly based on weighted categories
On-site or in-plant training	O*NET database	Exactly one value, selected randomly based on weighted categories
On-the-job training	O*NET database	Exactly one value, selected randomly based on weighted categories
Apprenticeship	O*NET database	Exactly one value, selected randomly based on weighted categories
Company name for second-to-last job	Mistral NeMo	Exactly one value, generated by the language model
City and state of employment	Mistral NeMo	Exactly one value, generated by the language model
Dates of employment	Mistral NeMo	Exactly one value, generated by the language model
Responsibilities for second-to-last job	Mistral NeMo	Exactly one value, generated by the language model

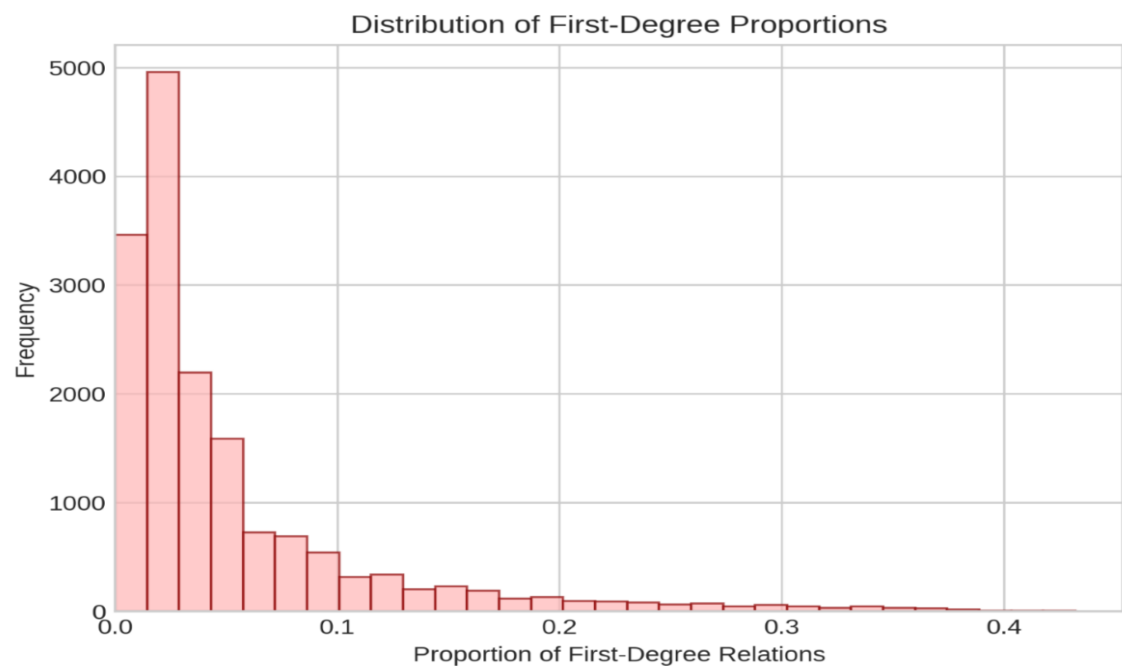
Appendix F: Additional graphs with results

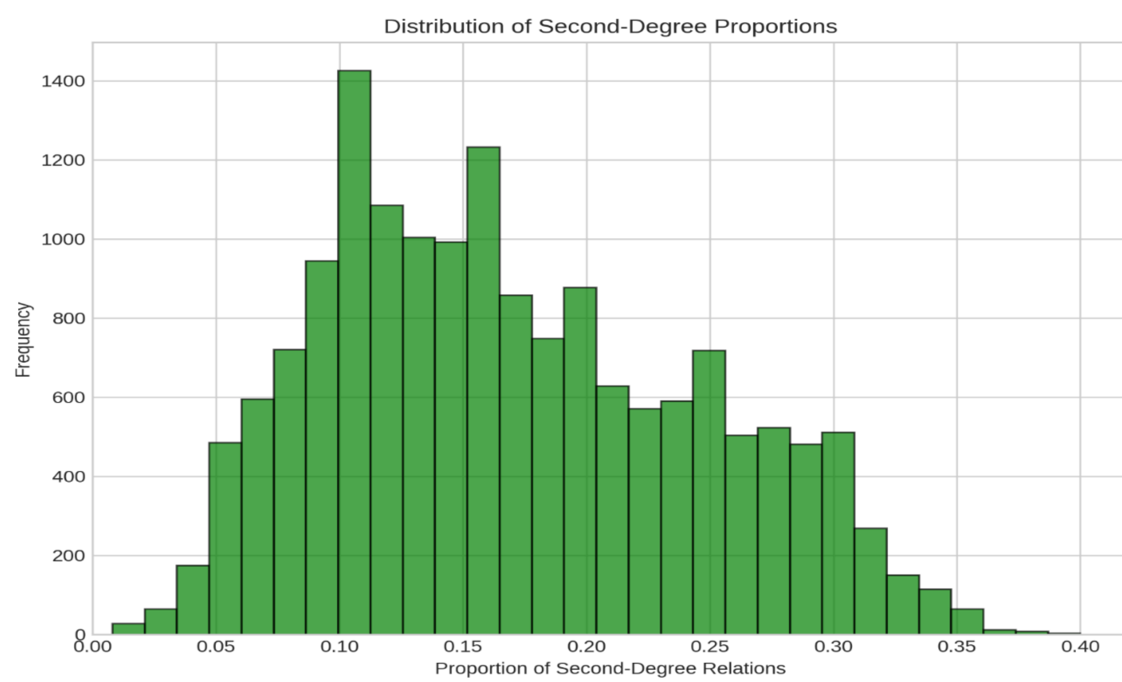


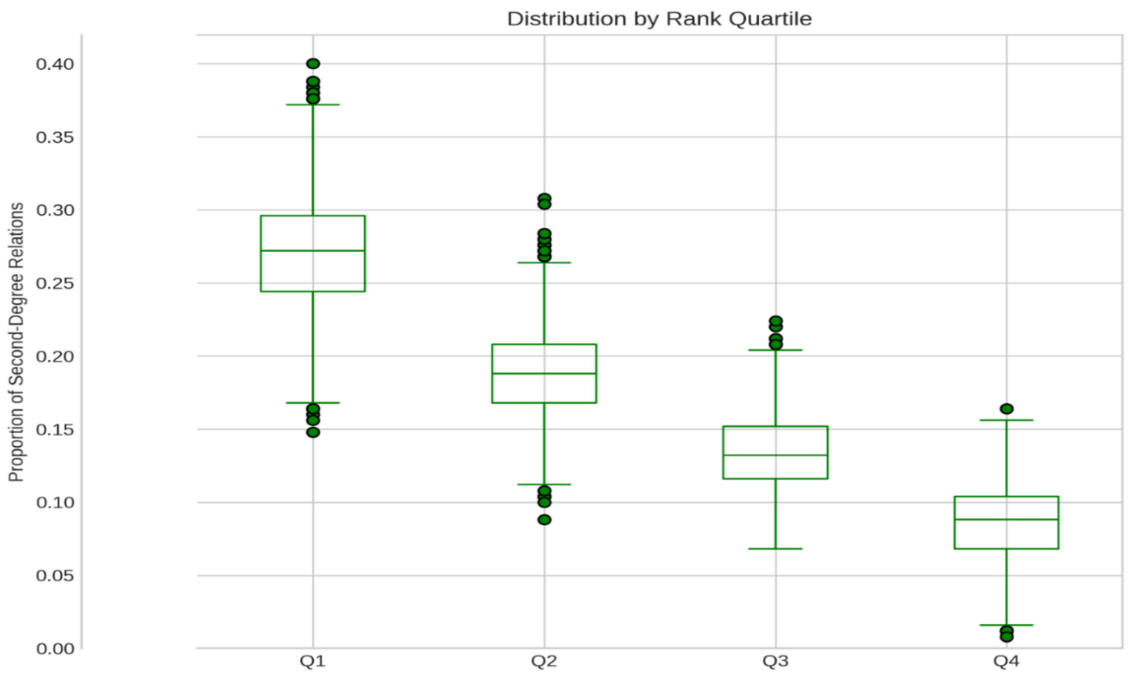


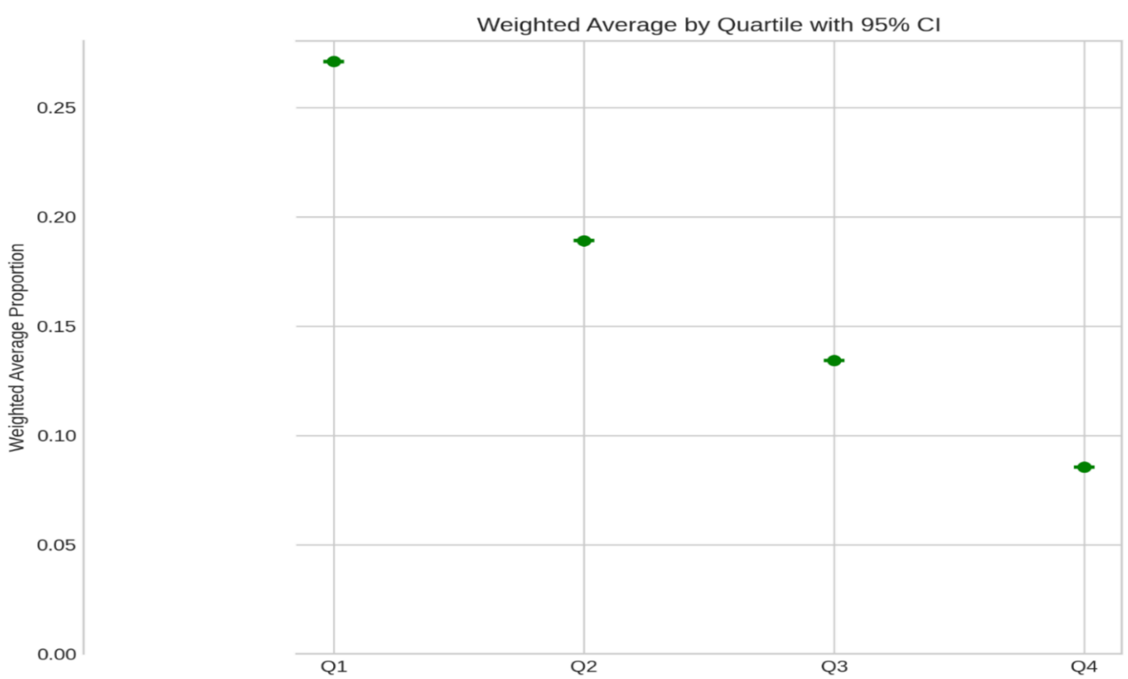

Weighted Average by Quartile with 95% CI
Weighted Average Proportion
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Q1
Q2
Q3
Q4

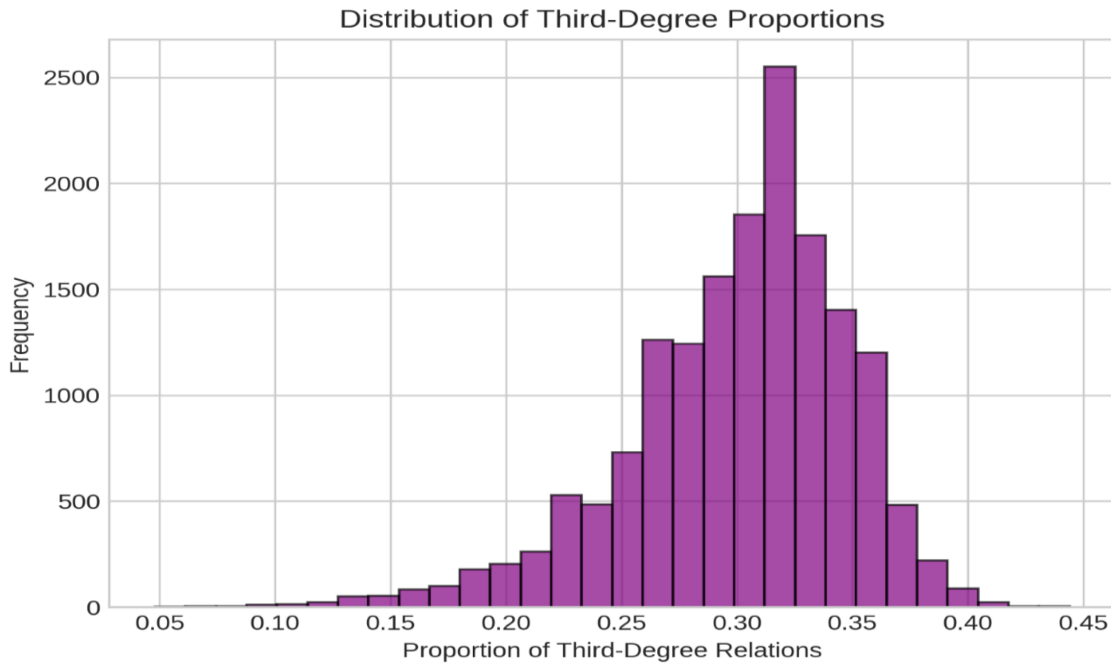

Guaranteed Task Performance by Rank Quartile
Proportion of Guaranteed Task Performance
1.0
0.8
0.6
0.4
0.2
0.0
Q1
Q2
Q3
Q4

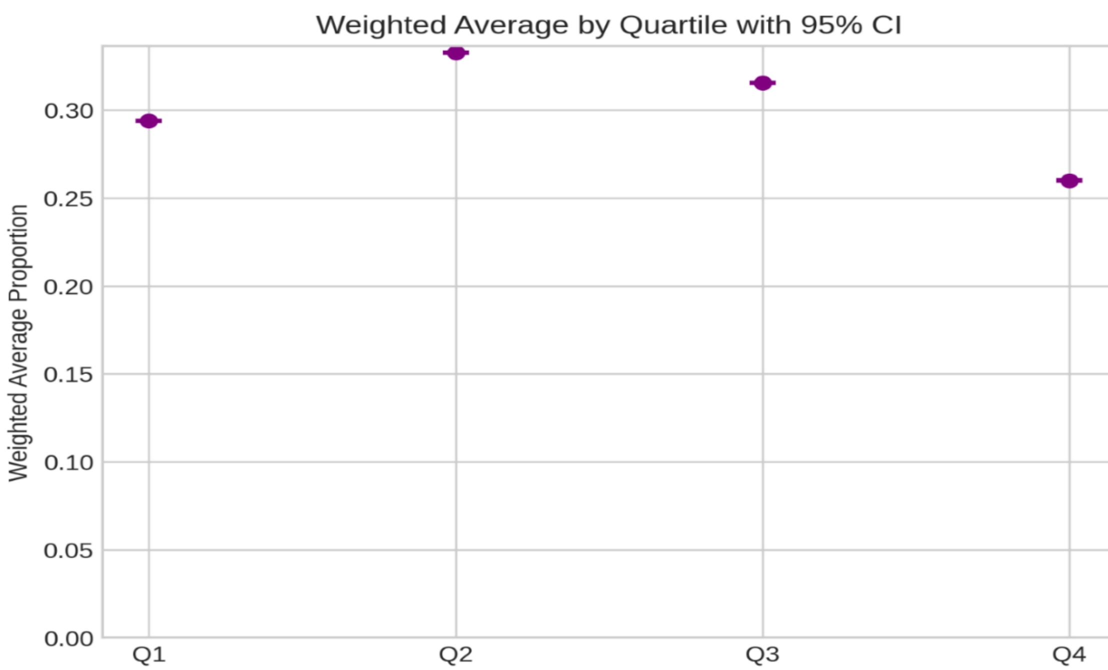

Distribution of First-Degree Proportions
Frequency
5000
4000
3000
2000
1000
0
0.0
0.1
0.2
0.3
0.4
Proportion of First-Degree Relations

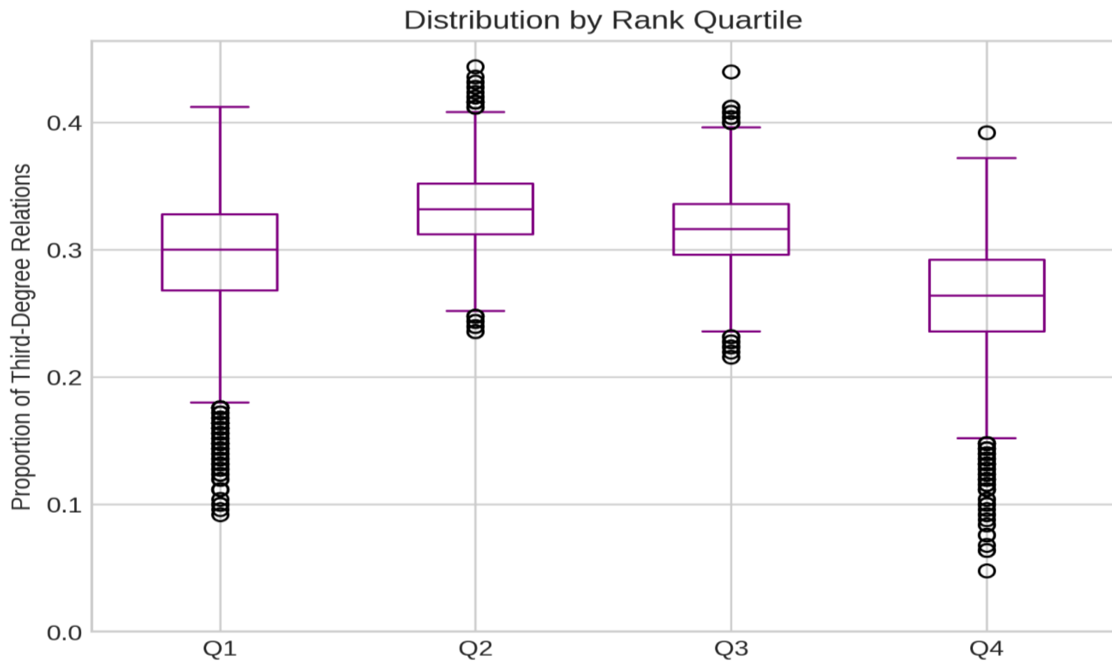

Distribution of Second-Degree Proportions
Frequency
1400
1200
1000
800
600
400
200
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Proportion of Second-Degree Relations

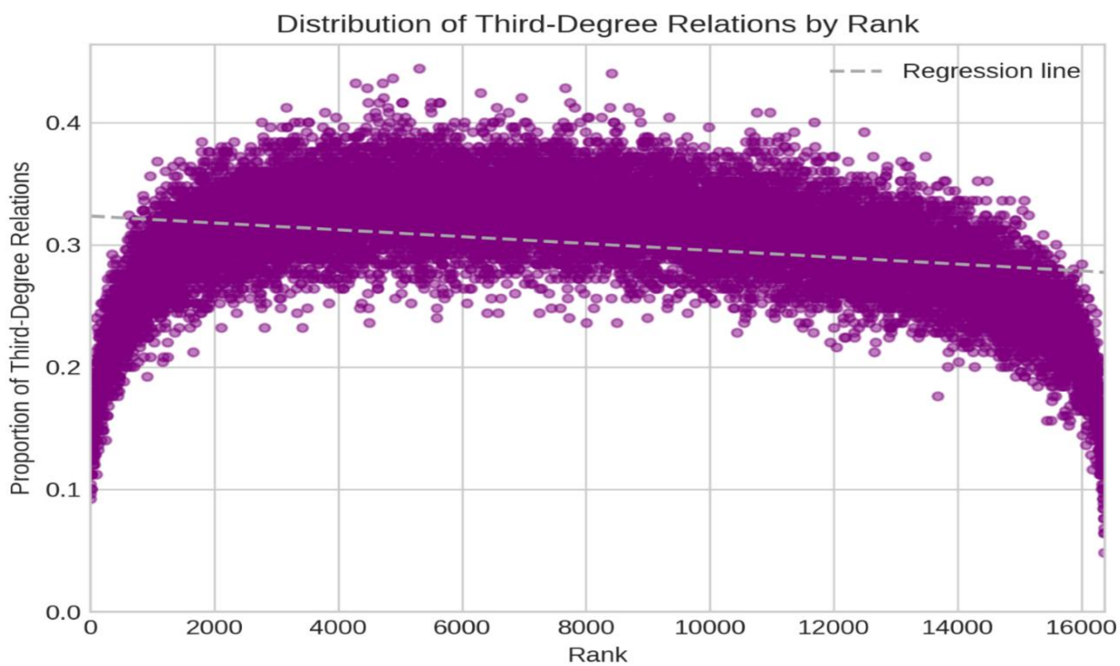

Distribution by Rank Quartile
Proportion of Second-Degree Relations
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Q1
Q2
Q3
Q4


Weighted Average by Quartile with 95% CI
Weighted Average Proportion
0.25
0.20
0.15
0.10
0.05
0.00
Q1
Q2
Q3
Q4


Distribution of Third-Degree Proportions
Frequency
2500
2000
1500
1000
500
0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
Proportion of Third-Degree Relations


Weighted Average by Quartile with 95% CI
Weighted Average Proportion
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Q1
Q2
Q3
Q4


Distribution by Rank Quartile
Proportion of Third-Degree Relations
0.0
0.1
0.2
0.3
0.4
Q1
Q2
Q3
Q4


Distribution of Third-Degree Relations by Rank
Regression line
Proportion of Third-Degree Relations
0.0
0.1
0.2
0.3
0.4
0
2000
4000
6000
8000
10000
12000
14000
16000
Rank