Timm Eichenberg, Nils Pudill, Britta Rüschoff, Anne Stockem Novo, Michael Vogelsang

Chapter 9
Developing Personas of Ideal-type Candidates in AI-related Jobs

An Exploratory Study Based on the Analysis of Online Job Postings

Abstract: This chapter presents a data-driven exploratory approach for identifying ideal-type candidate personas in the field of artificial intelligence (AI) based on the analysis of online job postings. The data used for analysis was collected from online job platforms for the German labour market. Latent Dirichlet Allocation (LDA) was used for topic modelling. A meaningful LDA solution with four topics was obtained, which formed the basis for identifying four different ideal-type personas that are sought for in AI-related job postings. The chapter provides an interpretation of the personas and their skill profiles, expertise and accountability level. The results can help job seekers, recruiters, and policymakers to better understand the skills and experiences that are sought by employers in the field of AI.

Keywords: personas, artificial intelligence, online job postings, profiles, skill, tasks

9.1 Introduction

Artificial intelligence (AI) changes business models, organisations and jobs. After years of research and development, today the diffusion of AI is driven by economic parameters. Workers who mainly perform routine tasks were the first to be affected by computerisation (Autor, Levy & Murnane, 2003). This assessment may change as generative AI affects the jobs of journalists, translators and artists. Also, the tasks of

Acknowledgements: We thank Tom Staab for his valuable contributions in writing and applying the Python code for the analysis.

Timm Eichenberg, Hochschule Weserbergland, Hameln, e-mail: eichenberg@hsw-hameln.de
Nils Pudill, Ruhr West University of Applied Sciences, Mülheim a.d.R., e-mail: nils.pudill@hs-ruhrwest.de
Britta Rüschoff, FOM Hochschule für Oekonomie und Management, Düsseldorf/Münster, e-mail: britta.rueschoff@fom.de
Anne Stockem Novo, Ruhr West University of Applied Sciences, Mülheim a.d.R., e-mail: anne.stockem-novo@hs-ruhrwest.de
Michael Vogelsang, Ruhr West University of Applied Sciences, Mülheim a.d.R., e-mail: michael.vogelsang@hs-ruhrwest.de

https://doi.org/10.1515/9783110790320-009

computer scientists will change as code is becoming a commodity which can be produced automatically.

Businesses are developing and implementing AI technology to support their activities or make AI technology the core of their business model itself. As a result, companies' labour demands in the field of AI are increasing. Therefore, we are interested in the profiles and skills which are looked for in the field of AI. First, we use a data-driven approach for identifying candidate personas in the field of AI based on online job postings. Second, we provide an interpretation of the personas and their skill profiles, expertise and accountability level. The results can help job seekers, recruiters, and policymakers to better understand the skills and experiences that are sought by employers in the field of AI.

9.2 Skills and Jobs Research

9.2.1 Skill-Biased Technological Change

AI is one of the most recent examples for technological change. On a macroeconomic level, the consequences of technological change are discussed intensively. Skill-biased technological change (SBTC) models point out that technological change increases the productivity and hence demand of skilled workers. The approach postulates that the skill premium, the wage gap between skilled and unskilled workers, increases depending on the supply of skilled workers and the elasticity of substitution between skilled and unskilled work. However, the empirical evidence between skill-biased technological change and wage inequality was "surprisingly weak" in the past (Card & DiNardo, 2002), p. 776). Subsequently, the focus has shifted to the differences between routine and nonroutine tasks (Autor, Levy & Murnane, 2003) and the polarisation thesis that the employment shares for high-paid persons and for low-paid personal service workers are growing. Overviews are provided for example by Hornstein, Krusell and Violante (2005) and Aghion et al. (2022).

Most recently, Acemoglu and Restrepo (2022) expanded the classical SBTC model by a task-based approach. The elasticity of substitution between skilled and unskilled labour becomes endogenous and depends on the characteristics of new tasks. They also take into consideration AI. They interpret AI as a technological platform that can be used for automation or leads to new labour-complementary tasks. As new tasks can foster skilled or unskilled work, the skill premium also depends on the type of new tasks created by AI (Acemoglu & Restrepo, 2022). Empirically, Acemoglu et al. (2022) show that the AI exposures of establishments are associated with negative and positive changes in the skill set required.

9.2.2 AI Jobs Research

Several authors argue that AI jobs require general or soft skills, such as problem-solving, creativity and teamwork (Squicciarini & Nachtigall, 2021) or social and emotional skills (Samek, Squicciarini & Cammeraat, 2021). The content of AI-related jobs is hence becoming more diverse (Verma, Lamsai & Verma, 2022). Samek, Squicciarini and Cammeraat (2021) further argue that there are complementarities between different occupational categories (managers, AI professionals, technicians) as recruitment of those often happens in the same organisation at the same time. Today, business knowledge is as important as technical skills for working in business intelligence and big data projects (Debortoli, Müller & vom Brocke, 2014), which is further supported by the emergence of business clusters such as “Data Science & Engineering,” “Software Engineering & Development” and “Business Development and Sales” (Anton, Behne & Teuteberg, 2020).

9.2.3 Methodological Approaches to the Analysis of AI-related Job Postings

Online job advertisements have been used as empirical databases at least since the year 2000, when Koong, Lui and Lui (2002) identified and ranked job postings in the field of information technology. Data are derived from several online platforms such as monster.com (Debortoli, Müller & vom Brocke, 2014) and hotjobs.com (Koong, Lui & Lui, 2002), indeed.com (Verma, Lamsai & Verma, 2022) or aggregators such as Burning Glass Technologies (BGT) (Acemoglu et al., 2022; Squicciarini & Nachtigall, 2021). When analysing and clustering the data, a broad range of statistical and specific Natural Language Processing (NLP) methods such as k-means (Litecky et al., 2010), Latent Semantic Analysis (LSA) (Debortoli, Müller & vom Brocke, 2014) or matrix factorisation (Anton, Behne & Teuteberg, 2020) is used. Ternikov (2022) provides a more recent overview of the methodology employed in IT job advertisement analysis and points to the fact that the approach used varies with the sample size of the data available.

In this chapter, Latent Dirichlet Allocation (LDA) is used, an approach which is applied for topic modelling. LDA is also used by Gurcan and Cagiltay (2019) who identify 48 skill topics, and by De Mauro et al. (2018), who determine nine skill sets and map them to four “Big data job families.”

9.3 Methodology

This study uses the concept of personas to comprehensively summarise the results generated from the job postings regarding the aspired candidates’ professional profiles along with the tasks and activities of the advertised jobs. The concept of a persona originated in human–computer interface design with the intent to facilitate user-centred designs by generating typical user profiles (Cooper, 1999). In its original application, personas were used to guide design processes by providing designers with evidence-based lifelike characters of users generated from available data on the core features (e.g., needs, goals, limitations, predicted behaviour) of the user. Since then, the use of the concept has expanded to areas outside human–computer interface design and today, a persona often generally refers to a fictional yet lifelike character that bundles the core features of, for example, a typical user (Nielsen, 2019), customer (Lehnert, Goupil & Brand, 2021) or, in this case, the targeted candidate. The approach taken in this study hence deviates from the original goal of the concept as a point of reference for design processes. Instead, we adopt the idea of a persona as a representation to identify the typical aspired candidate for a job in the area of AI, whose features do not necessarily have to be united in its entirety in one particular person.

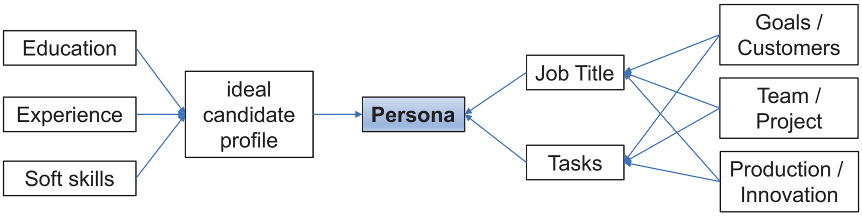

Figure 9.1: Building blocks of personas.

Figure 9.1 illustrates the structure of our approach: A persona is defined by characteristics of the aspired candidate (left side of Figure 9.1) and by characteristics of the advertised job (right side of Figure 9.1). In terms of aspired candidate characteristics, defining criteria are education, experience and soft skills. These aspired candidate characteristics are mapped to the job title and the tasks described in the job postings which are determined by the goals and customer orientation of advertising employer, its internal organisation and team structures, as well as innovation and production processes. With this persona-oriented approach, we expand the primarily skill-based view of related research and apply a more holistic view on the applicants targeted in online job postings.

9.4 Empirical Analysis

9.4.1 Data

The data subjected to analysis have been collected from three online job platforms for the German labour market from March to September 2022. The search term was restricted to “Künstliche Intelligenz” [engl.: Artificial Intelligence], yielding results for job postings with the keywords “Artificial Intelligence” in German and English language.

The following pre-processing steps were conducted: (1) Checking for duplicates by applying a cosine similarity measure and deleting duplicate entries, yielding in 8152 jobs postings. (2) Separating the job postings into job title and job profile, including degrees, and tasks. Depending on the data source, the separation of data could take place by considering structural elements such as HTML tags or by using an ontology approach with keywords. The keywords were manually updated by data exploration. (3) Normalisation of the data, converting to lowercase and elimination of stop words. (4) Partial language translation, depending on the model applied. (5) Stemming of the words.

9.4.2 Model Selection

Following our exploratory research approach, we focus on unsupervised models of machine learning to identify clusters of job postings. For the model selection, we first assumed that a set of (ideal-type candidates) profiles and a set of tasks in the clusters identified are related to each other unambiguously. Under this assumption, we used a hierarchical cluster approach and matching criteria based on the frequency distribution in a contingency table. k-means clustering with $k = 2$ was employed in three loops, so that we received a tree structure of clusters in three levels. We used TF-IDF vectors as inputs for k-means clustering as well as a Sentence BERT-Networks (SBERT) transformer called distiluse-base-multilingual pretrained model (Reimers & Gurevych, 2019). Fine tuning based on the corpus has been applied. With the multilingual SBERT embeddings, the language of the job postings should not matter. Interestingly, the results were similar to the first k-means loop with untranslated data. This suggests that in job postings written in English, specific tech-orientated skills are requested. In the next iterations, the matching criteria did not provide meaningful results, since the tasks in the job posting are too homogeneous.

We therefore deny the assumption of strongly separated clusters and use LDA for topic modelling (Blei, Ng & Jordan, 2003). The LDA algorithm assumes that words are characteristic for topics and topics are characteristic for documents. However, the topics cannot be observed directly in the documents, therefore they are latent. In this chapter, we use the LDA approach for classification of profiles, assuming that each latent topic represents an ideal-type persona. This assumption is in line with the former analysis, showing that profiles in the job postings are more heterogeneous than the tasks.

The topics themselves are described by a vector of words. The conceptual advantage of LDA over classification algorithms is that it allows for a distribution of words over the topics, for example a word can be part of more than one topic. Theoretically, the best allocation of the words to topics and topics to documents is found by maximising a joint probability function. A restriction of LDA is the assumption that the words are independent from each other. So the order of words does not influence the results (“bag-of-words assumption” (Blei, 2012)).

In this chapter, we apply the LDA function of the Gensim package for Python, which uses the variational inference approach for optimisation (Hoffman, Blei & Bach, 2010). This algorithm is very fast at finding a stable solution. Pivotal for the LDA algorithm is the parameter “number of topics.” We used an expert rating method to determine the most meaningful number of topics. We started with a solution containing a high number of topics (e.g., seven) and reduced the number of topics until a solution with a meaningful differentiation between the topics was reached. The LDA run with four topics delivered the most plausible solution. This is confirmed by the coherence scores. Generally, “coherence” of a topic model follows the definition that a set of statements is coherent if the facts within the statements support each other (Röder & Both, 2015). In this chapter, we used three different measuring approaches: UCI (Newman et al., 2010), UMASS (Mimno et al., 2011) and NPMI (Bouma, 2009). The first two are named after co-authors from the University of California and University of Massachusetts, respectively, and the third abbreviation stands for Normalised Pointwise Mutual Information. To calculate scores, we used implementations also provided by the Gensim topic modelling framework. The coherence scores C(K) for K topics show a steadily declining trend by nature. Using the “elbow-method” considering the inflection points of C(K), the UCI and UMASS approaches confirm the expert’s choice of four topics for the model (see Figure 9.2).

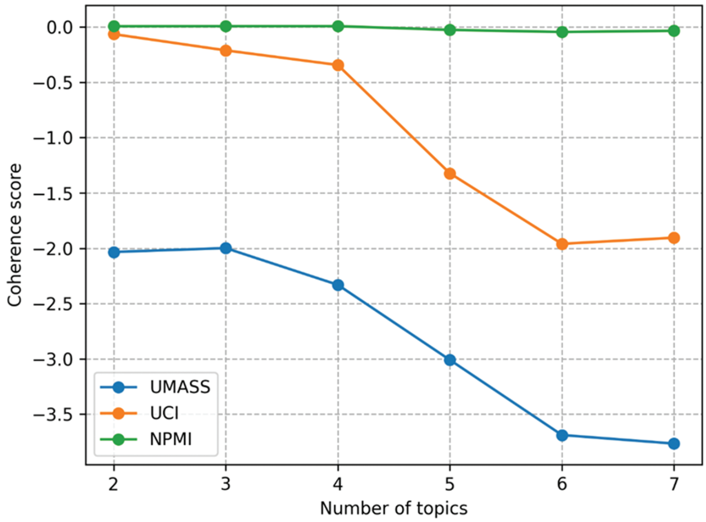

Figure 9.2: Coherence scores.

After topic modelling, we determined which postings are related to a topic and identified the tasks and job titles of these postings to describe the personas. The results are presented in the next section. Finally, we checked the results for robustness.

9.5 Results and Interpretation

9.5.1 Interpretation

Considering the model's output for four topics, characteristic keywords and the topics' positions in the coordinate system could be analysed. From the results, four different ideal-typical personas were derived, with each topic forming the basis for one persona.

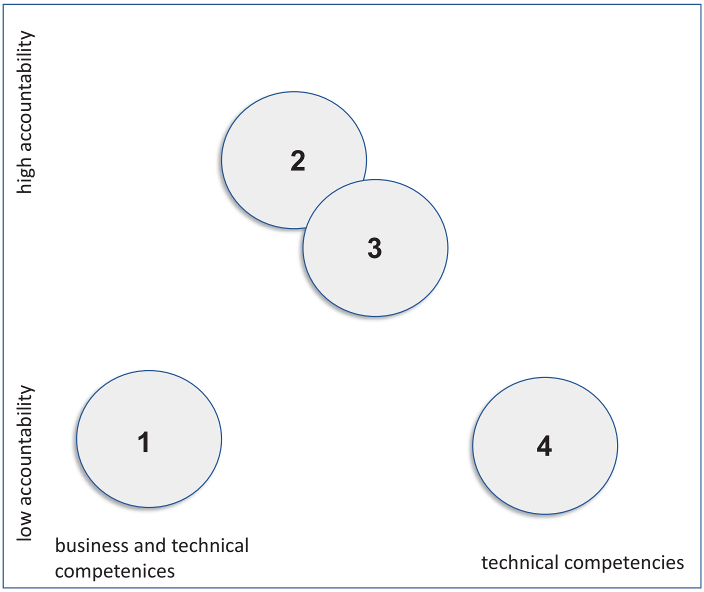

Figure 9.3: LDA model solution for four topics.

Figure 9.3 shows the stylised topics, based on the visualisation from the Python library pyLDAvis. The visualisation on two axes is to some extent arbitrary since the distribution of the topics in the coordinate system is multidimensional. In pyLDAvis, the axes (or principal components) are not specified beforehand but are generated based on a probability distribution of the words that are indicative of a certain topic. Therefore, Figure 9.3 serves entirely illustrative purposes. The meaning of the axes is likewise subject to interpretation and can only be construed conjointly with the interpretation of the four personas.

The horizontal axis can be interpreted as the level of expert competencies (left side = business and technical competencies; right side = technical competencies only). Business competencies are related to an education in business and economics [German: “Wirtschaftswissenschaften”] and/or managerial experience. The vertical axis can best be interpreted as overall accountability for results (bottom = low; top = high). Accountability refers to the person or team who is ultimately accountable for the success or failure of the outcome and who possesses the decision-making authority to approve or disapprove of the work completed by the responsible persons or teams.

9.5.2 Interpretation of the Four Personas

The proposed personas can be regarded as an explanatory approach for ideal-typical applicants in demand in the German labour market in the field of AI. In order to differentiate the four personas more efficiently, short designations for each persona were chosen. Each designation is based on the interpretation of the topics and commonly used generic IT job titles. The designations of the personas follow the interpretation of the results and do not necessarily have to be used in exactly the same wording in current or future job postings. The word clouds in this chapter show the words with the highest frequency per topic but not the most salient words. Therefore, the word clouds for the profiles show less differentiation than the LDA depicted.

9.5.2.1 Persona 1 – Junior Project Member

Persona 1 in the lower left corner of Figure 9.3 represents a candidate who is ideal for entry-level positions. Experiences [German: “Erfahrungen”] often refers to be “first experiences,” as became clear when analysing all sentences in the profiles containing the German word “Erfahrungen.” The fields of study of this persona can be business and economics [German: “Wirtschaftswissenschaften”] as well as computer sciences [German: “Informatik]. The interpretation of the tasks suggests involvement in development projects, mostly on a supporting level. We give this persona the designation “Junior Project Member.” The word clouds in figure 9.4 provide an illustration of the most frequent keywords for persona 1.

9.5.2.2 Persona 2 – Senior AI Manager

The second persona, located in the upper left corner of Figure 9.3, takes responsibility [German: “Verantwortung”] for customers. The desired educational background can be in business and economics or computer sciences. On average, the required work experience is longer than for persona 1 (which, however, cannot be seen in the word cloud). We label this persona as "Senior AI Manager." The word clouds in figure 9.5 provide an illustration of the most frequent keywords for persona 2.

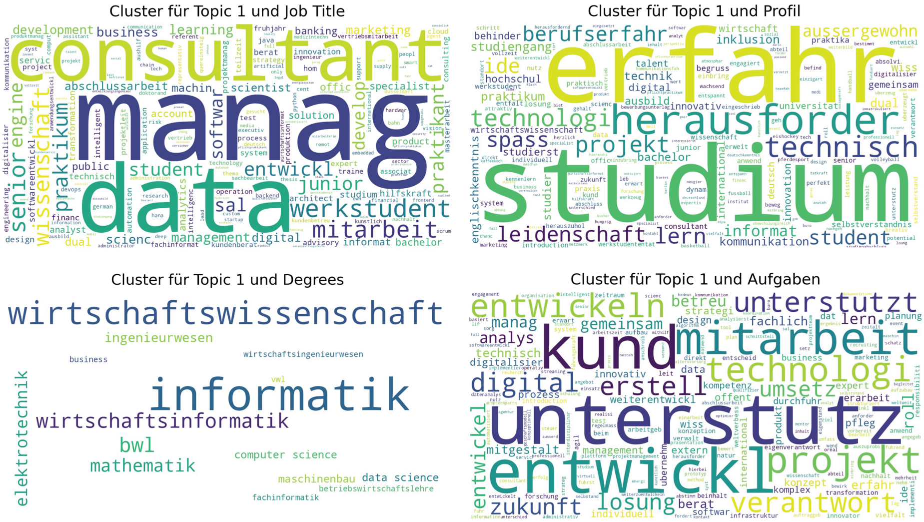

Figure 9.4: Most frequent keywords for persona 1.

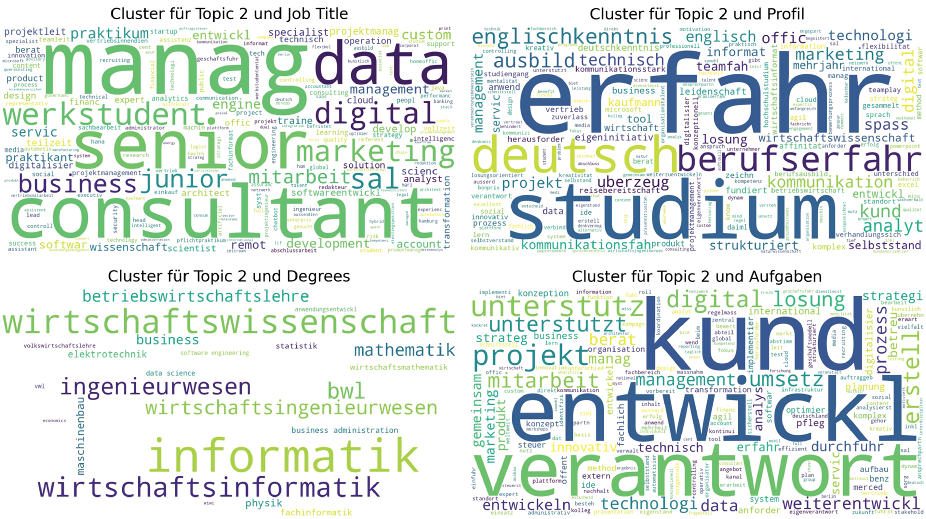

Figure 9.5: Most frequent key words for persona 2.

9.5.2.3 Persona 3 – Senior AI Architect

The third persona in the upper right corner of Figure 9.3 also takes responsibility for customers. But compared to persona 2, it focuses less on management tasks. The key terms associated with this persona highlight senior roles with a background in engineering and data science. The results concerning the profiles suggest that this persona has technical experience in data science and possesses competencies in technical domains, such as Python. In line with this, the degrees associated with this persona are primarily technical, emphasising information technology, computer science, mathematics and engineering. The tasks allocated to this persona suggest project-related responsibilities in the development of AI solutions, including consulting, again underlining the interpretation as a senior role with a strong technical focus. We give this persona the designation “Senior AI Architect.” The word clouds in figure 9.6 provide an illustration of the most frequent keywords for persona 3.

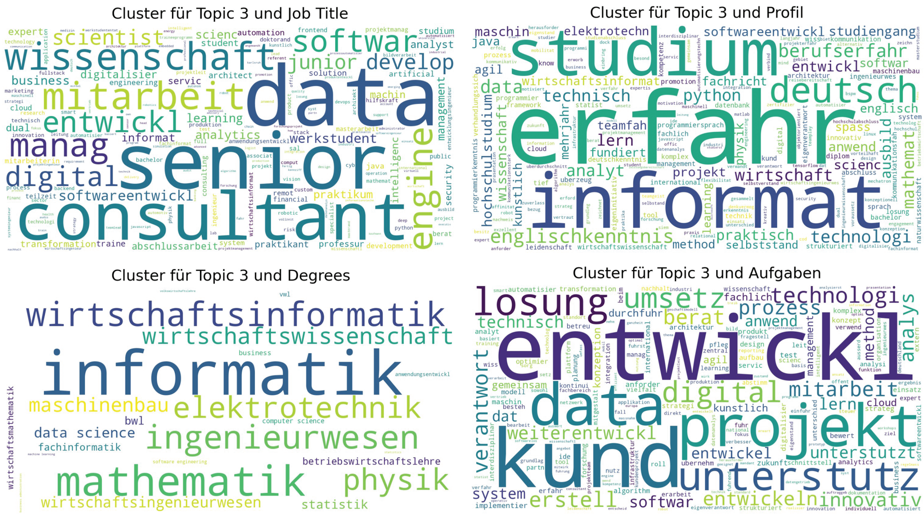

Figure 9.6: Most frequent keywords for persona 3.

9.5.2.4 Persona 4 – AI Developer

Persona 4 in the lower right corner of Figure 9.3 has a clear technical focus but is not involved in customer relationship management. The desired degrees are related to computer sciences and the tasks to be carried out are related to software development. A clear scientific background is necessary and desired experiences are related to specific programmes and frameworks. We describe this persona as “AI Developer.” The word clouds in figure 9.7 provide an illustration of the most frequent keywords for persona 4.

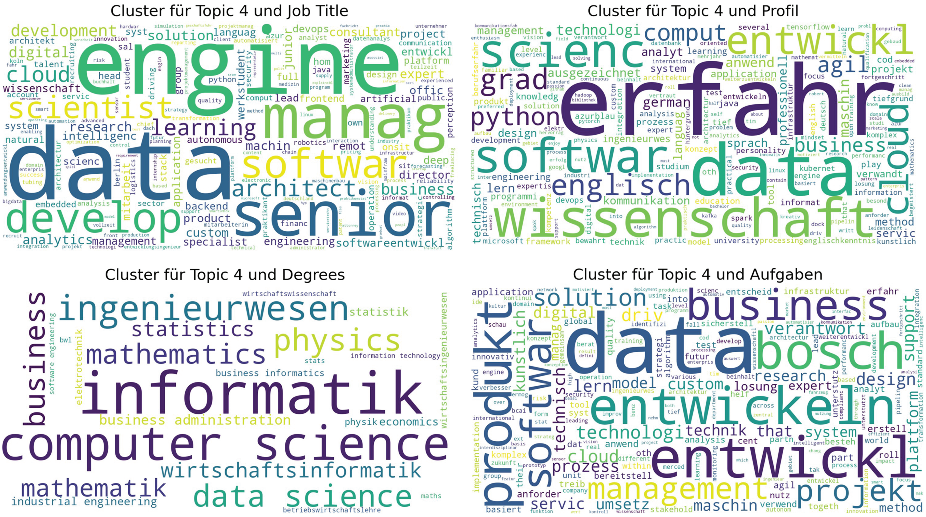

Figure 9.7: Most frequent keywords for persona 4.

9.5.3 Consolidation of the Results

The personas are developed from the interplay between desired profiles, job titles, tasks and further analyses such as an analysis of the context of the word “experience” [German: “Erfahrung”]. The consolidated results concerning the four personas are presented in Figure 9.8.

It seems that different personas are necessary to foster the diffusion of AI. The labour market, and in particular the growing field of AI-related jobs, hence provides opportunities for candidates with different backgrounds. Moreover, the personas do not necessarily have to be seen as separate individuals, but might also be interpreted as career-related development paths over time. Thus, a development from persona 1 to persona 2, 3 or 4 may occur as a person’s career path is unfolding: Starting from the junior position, a person having studied business or economics a person can become a Senior AI Manager later, provided that the necessary technical or statistical knowledge or experience is available. However, the possible career paths in AI jobs seem to be more diverse for persons who have studied computer sciences or related topics. They might become a Senior AI architect, an AI developer or a Senior AI Manager. However, any interpretation of the career paths at this point is speculative and does not emerge from the data itself.

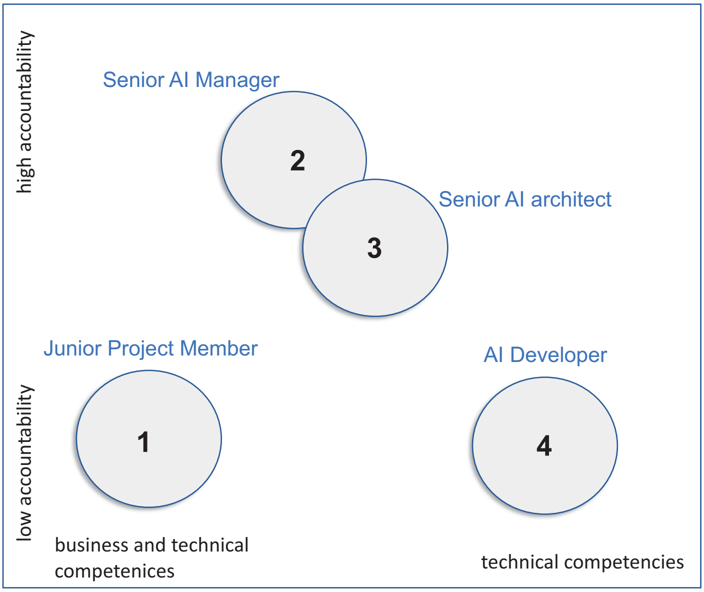

Figure 9.8: Interpretation of the LDA Model Solution for four personas.

9.5.4 Robustness

The LDA reacts sensibly to changes in the corpus. For a robustness check, we repeated the analysis ten times with 90% of the data used (i.e., 10% of the job postings were randomly removed in each of the ten runs), showing that the results of the LDA are to some extent arbitrary. As changes in individual words have a strong effect on the LDA algorithm, the four topic-solution presented before could not be replicated perfectly. In seven out of ten runs, the descriptions of a persona change slightly; in three out of ten runs, the generated topics do not match the personas described in this chapter. We expect that analyses with a larger database will generate more robust results.

9.6 Discussion

This study offers insights into the types of candidates and personas that are currently being recruited in the field of AI. This may not only add value for recruiting companies and current candidates but also help students in choosing their educational pathways. However, this study does not come without limitations. One limitation lies in the LDA and the assumption, that the words are independent from each other. This became obvious with the word “Erfahrungen” [English: Experience], where “Erste Erfahrungen” [English: first experiences] is related to a different persona than “langjährige Erfahrungen”

[English: long-term experiences]. Therefore, we checked the consistency with the persona descriptions manually. Another limitation lies in the adoption of a “naïve approach,” as we assume that each topic also represents an ideal profile. Therefore, we also employed a DBSCAN cluster algorithm after the LDA, but this combination did not improve the results. Future work should encompass these shortcomings, using larger databases and word embeddings, as used in the Top2Vec algorithm (Angelov, 2020). Also, it might be expected that the new generative pre-trained transformers (GPT) models from OpenAI (Brown et al., 2020; OpenAI, 2023), Google (Vaswani et al., 2017) and others also lead to new approaches for topic modelling and classification. Also, for a comparison of countries with different languages over time, GPT embeddings could be employed.

In addition to the practical value of the results to companies, candidates, and researchers interested in the impact of AI on the labour market, this study also offers methodological insights into the usefulness of the concept of personas. A (candidate) persona is a hypothetical person who represents the desired candidate (Rippler, 2022). The concept was adopted to facilitate the presentation of the ideal-type candidates, as it reflects an intuitive and clustered interpretation of the candidates being sought. When used in development processes, personas are deliberately created by organisations to reflect the optimal candidates a company is looking for. The personas are hence created based on organisational requirements. In this study, the concept of a persona is used to retrospectively draw a comprehensive conclusion about ideal-type candidates based on the requirements identified in various job postings. The derived personas hence do not represent requirements of a specific organisation but at labour market level. The application of personas in this study has proven to be a useful construct for presenting the diverse results regarding ideal-type candidates in a concise and intelligible manner. The approach taken in this study is therefore considered to be promising and may be applicable in future research to identify profiles that are sought for in the labour market and, possibly, how these profiles change over time.

9.7 Summary

From an explorative analysis of online job postings for AI jobs in Germany, four personas could be derived. These personas are designated as Junior Project Member (1), Senior AI Manager (2), Senior AI Architect (3) and AI Developer (4). The personas are differentiated according to their accountability and the required competencies. The chosen designations are based on experts’ interpretation and serve as a suggestion for generic labels. They do not necessarily correspond with the various job titles used by companies in their job postings in the underlying data set. Assuming that different personas might be necessary for a successful AI-related project, the results can be interpreted as a hint for a project-based diffusion of AI in companies.

Bibliography

Acemoglu, D., Autor, D. H., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, *40*(S1), 293–340. https://doi.org/10.1086/718327.

Acemoglu, D., & Restrepo, P. (2022). A task-based approach to inequality. IFS Deaton Review of Inequalities, March 2022. https://ifs.org.uk/inequality/wp-content/uploads/2022/03/A-task-based-approach-to-inequality-IFS-Deaton-Review-of-Inequalities-1.pdf (accessed 9 July 2023).

Aghion, P., Antonin, C., Bunel, S., & Jaravel, X. (2022). The effects of automation on labor demand: A survey of the recent literature. CEPR Discussion Paper Series (DP16868). https://repec.cepr.org/repec/cpr/ceprdp/DP16868.pdf (accessed 9 July 2023).

Angelov, D. (2020). Top2Vec: Distributed representations of topics. https://arxiv.org/pdf/2008.09470.pdf (accessed 9 July 2023).

Anton, E., Behne, A., & Teuteberg, F. (2020). The humans behind Artificial Intelligence – An operationalisation of AI competencies. ECIS 2020 Proceedings. Research Papers. https://aisel.aisnet.org/ecis2020_rp/141 (accessed 9 July 2023).

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801.

Blei, D. M. (2012). Probabilistic topic models. *Communication of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bouma, G. (2009). Normalized (Pointwise) Mutual information in collocation extraction. Proceedings of Gesellschaft Für Sprachtechnologie Und Computerlinguistik (GSCL), 30, 31–40. https://svn.spraakdata.gu.se/repos/gerlof/pub/www/Docs/npmi-pfd.pdf (accessed 9 July 2023).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, S., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165.

Card, D. E., & DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics*, *20*(4), 733–783. https://doi.org/10.1086/342055.

Cooper, A. (1999). The inmates are running the asylum. In: U. Arend, E. Eberleh, & K. Pitschke (Eds.), *Software-Ergonomie '99: Design von Informationswelten*. Vieweg & Teubner Verlag. https://doi.org/10.1007/978-3-322-99786-9_1.

Debortoli, S., Müller, O., & vom Brocke, J. (2014). Comparing business intelligence and big data skills: A text mining study using job advertisements. *Business & Information Systems Engineering*, *6*(5), 289–300. https://doi.org/10.1007/s12599-014-0344-2.

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing and Management*, *54*(5), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004.

Gurcan, F., & Cagiltay, N. E. (2019). Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling. *IEEE Access*, *7*, https://doi.org/10.1109/ACCESS.2019.2924075.

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online learning for Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, *23*, 856–864. https://papers.neurips.cc/paper/2010/file/71f6278d140af599e06ad9bf1ba03cb0-Paper.pdf (accessed 9 July 2023).

Hornstein, A., Krusell, P., & Violante, G. L. (2005). The effects of technical change on labor market inequalities. In: P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1B, 1275–1370). Elsevier.

Koong, K. S., Lui, L. C., & Lui, X. (2002). A study of the demand for information technology professionals in selected internet job portals. *Journal of Information Systems Education*, *13*(1), 21–28. http://jise.org/Volume13/n1/JISEv13n1p21.pdf.

Lehnert, K., Goupil, S., & Brand, P. (2021). Content and the customer: Inbound ad strategies gain traction. *Journal of Business Strategy*, *42*(1), 3–12. https://doi.org/10.1108/JBS-12-2019-0243.

Litecky, C., Aken, A., Ahmad, A., & Nelson, H. J. (2010). Mining for computing jobs. *IEEE Software*, *27*(1), 78–85. https://doi.org/10.1109/MS.2009.150.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing Semantic Coherence in Topic Models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 262–272, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Newman, D., Han Lau, J., Grieser, K., & Baldwin, T. (2010). Automatic Evaluation of Topic Coherence. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 100–108, Los Angeles, California. Association for Computational Linguistics.

Nielsen, L. (2019). *Personas – User Focused Design*. Springer. https://doi.org/10.1007/978-1-4471-7427-1

OpenAI. (2023). GPT-4 technical report. https://doi.org/10.48550/arXiv.2303.08774 (accessed 9 July 2023).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. https://doi.org/10.48550/arXiv.1908.10084 (accessed 9 July 2023).

Rippler, S. (2022). *Das Persona-Prinzip. Erfolgreiches Recruiting mit Candidate Personas*. Springer Gabler. https://doi.org/10.1007/978-3-658-38979-6.

Röder, M., & Both, A. (2015). Exploring the space of topic coherence measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, 399–408. http://doi.org/10.1145/2684822.2685324

Samek, L., Squicciarini, M., & Cammeraat, E. (2021). The human capital behind AI: Jobs and skills demand from online job postings. *OECD Science, Technology and Industry Policy Papers*, 120. https://doi.org/10.1787/2e278150-en.

Squicciarini, M., & Nachtigall, H. (2021). Demand for AI skills in jobs: Evidence from online job postings. *OECD Science, Technology and Industry Working Papers*, 2021/03. https://doi.org/10.1787/3ed32d94-en.

Ternikov, A. (2022). Soft and hard skills identification: insights from IT job advertisements in the CIS region. *PeerJ Computer Sciences*, 8. http://doi.org/10.7717/peerj-cs.946.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://doi.org/10.48550/arXiv.1706.03762.

Verma, A., Lamsai, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education*, *36*(1), 63–73. https://doi.org/10.1177/0950422221990990.