Mitigating Bias in Recruitment: A Practical Approach to CV De-identification Considering Privacy Sensitive Information

Sascha Löbner[1] [0000-0001-9164-1919], Jetzabel Serna[2] [0000-0003-4515-222X], Frédéric Tronnier[1] [0000-0002-6202-4871], Welderufael Tesfay[1] [0000-0002-1087-2019], and Kai Rannenberg[1] [0009-0000-6142-7161]

[1] Goethe University, Frankfurt am Main, Germany
{sascha.loebner,frederic.tronnier,welderufael.tesfay,kai.rannenberg}@m-chair.de
[2] Cyber4People GmbH, Germany
jetzabel.serna@cyber4people.com

Abstract. Curriculum Vitae (CVs) often contain sensitive personal information, which poses risks of discriminatory use by employers during the hiring process. To address this, there have been various research approaches to de-bias the various phases of the hiring process. However, these initiatives often fall short of resolving the core issues effectively. In this paper, we propose a process and proof of concept aimed at de-identifying Privacy Sensitive Information (PSIs) in CVs. We run rigorous evaluations to de-identify relevant PSI outlined in the EU GDPR (e.g., art. 9) and synthesize findings from state-of-the-art research on text and CV de-identification. Our designed prototype is able to firstly identify PSI specific to CVs and then de-identifies these PSI within the Europass CV structure. This novel approach demonstrates a promising path towards promoting fairness and preserving privacy in the recruitment process.

Keywords: De-identification, PSI, resume, CV, privacy, fairness.

1 Introduction

In recent years, technological advances and digitization efforts have significantly transformed traditional recruitment processes. The integration of data-driven tools such as Automated Hiring Systems (AHS) promise to streamline operations, enhance efficiency and mitigate biases which are inherent in conventional hiring methods [19]. However, despite the widespread adoption of such technologies and claims of neutrality and objectivity, recent studies [32] reiterate that recruitment processes remain vulnerable to human biases and prejudices, influencing the decision-making at various stages - from job posting and initial screening to interview evaluations. This susceptibility can lead to disparities in candidate selection and perpetuate inequalities in employment opportunities [22]. Since CVs contain many personally identifiable and Privacy Sensitive Information (PSI) [38], such as name, sex, age, ethnicity, AHS, and decision makers often fail to focus solely on candidates' qualifications and experiences. Furthermore, recruiting practices often involve outsourcing and multiple stakeholder (e.g., headhunting or job platforms). Thus, handling vast amounts of job applications loaded with PSI raises privacy concerns, and increases the risk of bias and non-compliance with, e.g., the General Data Protection Regulation (GDPR) and AI Act. In this context, de-identification of CV data could enhance the objectivity and impartiality of the employment procedures, by anonymizing or pseudonymizing relevant PSI in CVs. Data-driven tools and decision makers could then focus solely on the candidates' core qualifications, skills, and experiences, thereby fostering a fair and merit-based selection process. More precisely, CV de-identification helps mitigate bias and discrimination in the initial screening phase [22] by providing unbiased training datasets for the training of the AHS system model and reducing unconscious bias in cases where human intervention is involved. De-identification also minimizes privacy risks associated with handling (i.e., storing, sharing, and transferring) large volumes of CVs containing PSI. In the context of outsourcing, it significantly reduces the risk of inadvertently sharing personal data with third parties and potential data breach exposures. Additionally, de-identification enhances transparency by demonstrating the organization's commitment to fair and unbiased recruitment. Actively promotes diversity, inclusion, and equal opportunities, particularly for underrepresented groups, ensuring that initial screening is solely based on merits and qualifications. Ultimately, these measures support compliance with anti-discrimination laws and help mitigate liability related to discriminatory hiring practices.

This work contributes a practical road map for effective CV de-identification to systematically address the critical issue of PSI exposure in CVs, which in turn may impact fairness and reduce bias in the candidate screening phase during recruitment. Our contributions can be summarized as follows:

1. We identify key categories of PSI often present in CVs and relevant for de-identification with the aim to mitigate potential bias during job screening.
2. We propose various methods to locate and de-identify these PSI types usually present in CVs.
3. We discuss the implications and challenges of de-identifying CVs and provide implementation considerations based on a prototype.

2 Related Literature

In this section, we explore research on fairness and bias in hiring practices, the use of AI in recruitment, techniques for detecting and de-identifying PSI in CVs, and the role of synthetic data in privacy preservation.

Studies on fairness in hiring investigate data bias issues in AI-supported recruitment and related ethical implications. Raghavan et al. [32] show the interplay between algorithmic de-biasing and anti-discrimination law, while Gu et al. [13] analyze how race, gender, and qualifications impact fair judgment in hiring. Rigotti and Fosch-Villaronga [34] define and operationalize fairness in AI recruitment, to prevent data protection violations and social discrimination.

Tronnier et al. [40] categorize biases and ethical issues in digital services, discussing necessary stakeholder trade-offs. Techniques to mitigate bias include algorithm fine-tuning and anonymity in resumes. Deshpande et al. [7] present a fair matching algorithm that analyzes demographic factors using cosine similarity. Kang et al. [22] show that "resume whitening" helps minority groups, however, organizational diversity statements did not reduce discrimination. Parasurama and Sedoc [30] analyze a gender debiasing method in NLP that was found ineffective, highlighting challenges in fully anonymizing gender in resumes and limitations of lexical approaches. Using regular expressions, LSTM and BERT, Jensen et al. [19] explore the de-identification of job vacancies in stack-overflow focusing on the PSI location, organization, contact, name and profession. The detection and de-identification of privacy sensitive information (PSI) are critical in maintaining privacy. Ribeiro et al. [33] analyze automated de-identification methods in clinical textual data. They find that achieving a high privacy with recall of 100 % causes information loss and reduced readability. Tesfay et al. [37] introduce a privacy bot for detecting GDPR-related sensitive data, also included in this work. With the rise of Generative AI (GenAI) applications and services, concerns about data quality, misinformation and protection have been raised and studied [46, 42]. Recent academic research is tackling this topic from different dimensions, i.e., by studying how synthetic, GenAI-generated, data can be used as a surrogate for real data, to not contain private information. To this end, [12] provide a framework to quantify the three privacy risks of singling out, linkability, and inference risks in synthetic data. Zhao et al. [45] also study synthetic text data generation for privacy protection. The authors propose a solution that "combines the iteratively optimized mindset from genetic algorithms to align the distribution of synthetic text with that of private text." [45]. Upon evaluation, the solution maintains a high degree of utility while providing robust privacy levels. Similarly, [21] combine a generative autoencoder, federated learning, and differential privacy to generate synthetic data without revealing the original data. Utility and data protection outperform existing solutions for attribute classification tasks, and thus, show the worth of synthetic data in future use cases. Other authors focus on synthetic speech data in their analysis [20]. The authors develop a network to verify the authentication of data and identify forged segments in the data, demonstrating robust capabilities. Bias in Generative AI (GenAI) applications are also a point of study in academic research, i.e. with a systematic literature review on gender bias in AI provided in [41].

Concerning synthetic data, [3] provide a framework for the creation of synthetic datasets with different types of biases to study the interconnectedness of biases and possible mitigation mechanisms. An SoK on privacy-preserving data synthesis is provided by [16]. Skondras et al. [36] generated synthetic resumes with GPT. They use their generated data for NLP-based resume classification, achieving an accuracy for synthetic data of 85 %. This motivates the need of de-identification of synthetic CV data for resume classification scenarios. First work on resume de-identification using in-prompt anonymization is conducted in [26], whereby the authors find only limited loss in performance in their model. Schneider et al. [35]

study how generative AI applications can be trained to avoid the identification of created synthetic data. Regulatory concerns from the European AI Act and GDPR compliance have been discussed by several researchers [14, 39]. Memmert et al. [27] test the usability to aid of generative AI in design science research. Technological advancements also play a significant role in automated hiring processes. Zaroor et al. [44] study resume classification with real-world data, utilizing matching algorithms to maintain low error rates. Javed et al. [18] test a job role classification system using SVM-kNN, covering various taxonomies. Pal et al. [28] explore the use of Naive Bayes, Random Forest, and SVM in automated hiring, highlighting the shift towards electronic resources in recruitment.

To the best of our knowledge, there is currently no comprehensive guideline that covers all important PSI types and provides practical implications for the implementation of CV de-identification tools.

3 Methodology

In this study, we adopt a problem-centered approach to design science research [31, 9, 15]. We also take into consideration the work of [24] that provide a method and approach on how to employ synthetic data for use cases in academic research. Following these steps, we aim to produce a robust CV de-identification prototype that can be adapted as a starting point for organizational implementation.

Problem Identification and Motivation: Why CV de-identification is important might not be intuitive at a glance, since employers aim to collect as much information as possible about an applicant. But with the outsourcing of job application processes to job platforms and automated CV analysis, de-identification becomes essential to ensure compliance with data protection regulations (e.g. GDPR). Also for fairness and equality, PSI such as gender [30] or age [43] should be removed in CV analysis. This confronts companies with the burden to judge which PSI types should be de-identified. Thus, our approach promotes fairness in automated decision-making and contributes to fair and trustworthy use of AI.

Defining Objectives of a Solution: We set the following objectives for CV de-identification tools and demonstrate the main actions with our prototype:

Ensuring that no discriminatory PSI types are used within automated recruitment processes to effectively mitigate biases and comply with GDPR.

Maintaining the utility of CVs for job suitability assessment. This objective is in line with the privacy by design principle of full functionality [5].

Ensuring relevance of PSI de-identification. To guarantee that CV de-identification remains relevant for the standardized format across the EU.

Design and Development: To identify PSI types in CVs that may result in bias, we build on the comprehensive mapping of Chua et al. [6] who compared and combined different categorizations of PSI types from literature. We apply this framework similar to Löbner et al. [25] for the CV de-identification. The Europass

[10] CV template represents the EU's efforts to standardize the presentation of individuals' skills and qualifications.To meet our objective of relevance, we analyse its free text fields. We evaluate GDPR-compliance of identified PSI types, focusing on Articles 4 and 9, that deal with the protection of personal and sensitive data processing. To identify suitable PSI detection techniques, we compare different approaches to identify PSI in texts in general, following Tesfay et al. [38]. An overview of relevant NER techniques for PSI detection is provided by Löbner et al. [25]. Identifying a suitable PSI de-identification techniques includes balancing the privacy and data utility trade-off. ISO/IEC 20889:2018-11 [17] provide different de-identification techniques such as generalization, pseudonymization or anonymization that we evaluate for each relevant PSI type.

Demonstration: To run our prototype in a controlled environment without risking exposing real world private data we decided to use synthetically generated data from ChatGPT4 2023 similar to Skondras et al. [36]. We manually validated all generated CVs to ensure the synthetic data contained realistic, diverse and correct CVs types aligned with typical resume content. For the prototype, we implement a simple model based on BERT [8] and the regex module. We have chosen BERT due to its superior performance in similar NLP tasks [23].

Evaluation: The PSI detection prototype is evaluated with the quantitative approaches of F1-score, precision and recall. We assess qualitatively the readability and utility of all de-identified text by comparing them one by one with the original.

4 Results

4.1 Design and Development

Identifying biased PSI types in CVs: In this paragraph, we present potentially biased PSI types in CVs, describe why they are might be biased and evaluate whether they are covered in the GDPR. Chua et al. [6] define six PSI categories that are Lifestyle behavior, Social Economic, Tracking, Financial, Authenticating, Medical-health. Although rare cases might exist, we exclude financial information, medical information and authenticating information because these categories were not found in our CV samples. Thus, we focus on the remaining criteria that are social-economic, tracking data and lifestyle (including behavioral) data. In Table 1 we analyse the characteristics mentioned in [6]. Moreover, we provide examples and evaluate the probability of occurrence based on how common the category is to appear in CVs. In the following, we present each characteristic and provide an assessment of whether a characteristic should be de-identified. To provide a more comprehensive overview, we categorize characteristics in more detailed PSI types. If we conclude that a PSI type has a high occurrence, we add a PSI tag to it and explore its de-identification. However, PSI types with low occurrence can also identify a person or cause biases what will be part of our future work. Finally, we compare the elicited tags with the GDPR (see sec. 4.1).

Table 1. Classification of PSI detected, based on [6] including Occurence (Occ).

Category	Characteristics	CV Criteria	Occ.	Tag
Lifestyle-behaviour	Beliefs	Religious views, philosophical beliefs	Low	
	Preferences or interests	Hobbies	High	HOB
	Family / Relationship	Fam. structure, relationships, marriage	High	FAM
	Politics	Membership in parties, organizations	Low	ORG
Social-economic	Ethnicity	Ethnic origin, race, language	High	NATION
	Physical characteristics	Photos	High	
	Demographics	Name, date, age, date of birth, gender, sex, nationality	High	NATION, DATE, PER, PNOUN, GNOUN, TOA
	Professional career	Prev. employer, job roles/details, salary, education	High	ORG, EIN
Tracking	Contact information	email address, phone number, address	High	EMAIL, PHONE
	Location	Location	High	LOC
	Computer device details	URL	High	LINK

Lifestyle behaviour includes all information on an individual's lifestyle and behaviors that shape their relationships, community connections, habits, preferences, beliefs, or opinions [6]. **Beliefs**, e.g., *Religious or philosophical beliefs* (cf. art. 9 of the EU GDPR) can cause bias in CV classification. Nevertheless, expressing religious or philosophical beliefs in a CV is quite uncommon, which is the reason why we do not consider it a high occurrence. **References or interests**, e.g. *Hobbies*, generally do not reveal PSI relevant to hiring decisions. However, certain hobbies, e.g. competitive sports' achievements like, Olympic participation, can increase identifiability. Additionally, some hobbies are statistically associated with gender, potentially enabling indirect gender inference [30]. Given their limited relevance to assessing job qualifications and the risk of introducing bias or indirect discrimination, we excluded hobbies from the PSI detection scope to align with fair hiring practices and promote equal opportunity. **Politics**, e.g., *membership in organizations or parties* could reveal personal affiliations, thus de-identifying could prevent bias related to perceived ideologies. Although we have not assigned a specific tag, organization-membership is covered by ORG.

Social-economic includes all PSI types that describe a person's ethnicity, physical characteristic, demographics or professional career are clustered here. **Ethnicity**, e.g., prejudices due to *ethnic origin* are a major issue for fairness in job role assignment. While removing ethnicity is often effective in mitigating ethical bias, there are reasons, such as maintaining team diversity, why it is not the sole approach. We consider ethnic related issues as highly present and classify the occurrence as high. Even indirect identifiers can reveal ethnic origin [22]. **Language** is relevant information for the application assessment but can also reveal the ethnic origin [22]. Thus, de-identification approaches should be considered carefully. **Physical characteristics** such as *photos* in traditional applications are suspected of causing bias. However, Frauendorfer et al. [11] found that including a photo on the CV does not significantly change the personality assessment. Still, the qualification of a person does in general not depend on how beautiful they look. De-identification should not cause a loss in utility. Regarding **demographics**, several important PSI types exist that should be carefully considered. *Name* is a direct identifier and should be de-identified. *Date* is an indirect identifier which can result in identification of a person when coupled with other data [29], as such it should in general be de-identified. A variety of dates exist among a CV that might require different de-identification approaches to keep utility optimal. These can contain time spans, e.g. work experience, graduation dates or age. *Date of birth / age* is an indirect identifier and should as a date be considered for de-identification. Historically, age had a huge role in salary negotiation, especially for companies that implemented a seniority based salary system. Age discrimination is a common hiring issue Zaniboni et al. [43], that we thus aim to avoid. Especially in Europe, it is common sense that fair evaluation of candidates should be based solely on qualifications and not on *gender* roles. Thus, favouring female candidates just because of quotas, or favouring male candidates just because of their gender is in the following referred to as gender biases. With our approach we aim to avoid these biases despite the challenges that exist in finding a practical implementation [30]. Regarding *national background*, national prejudice exist that are likely to bias decisions. Since national background caused biases are also ethical bias, we decided to combine them in the tag NATION. Regarding **professional career**, *previous employer* might cause a bias due to the good or bad reputation of the employer. Thus, de-identification should be considered. *Job details* are a valuable source of the actual task and experience of an applicant, and should thus not be removed. *Job roles* also provide valuable insights about an employee and might answer the question about experience in leadership or qualification for certain responsibility levels. Although they should not be removed, job roles can reveal the applicant's gender, which should be considered within the PSI type gender. *Salary* often has to be provided as expectation and thus, should be preserved. *Education* can reveal university, year of graduation, age, gender, ethnicity and even socio-economic background. Thus, de-identification can help to focus only on qualifications.

Tracking includes all information that can be utilized to locate a person, which ultimately aids in identifying an individual. **Contact information** is not required when outsourcing the initial candidate screening. When removing personal contact information, it is easier to share CVs and be privacy compliant, as many stakeholders access such information. *Email addresses* often contain fragments of the name or age and should therefore be de-identified. *Phone numbers* can be used to call a person to reveal their name but also to validate recommendation letters. Thus, de-identification should be considered carefully. An *address* can be easily used to identify a person and should thus be de-identified. **Location** can be used to identify a person, especially in combination with other data and should thus be de-identified. **Computer device details**, e.g., *URLs* can point to the applicant's homepage or contain other PSI types and should be considered for de-identification.

GDPR comparison: The GDPR is designed to protect the privacy of online users by promoting privacy-by-design [5] and by default, and by prohibiting the

Table 2. Coverage of PSI types in the GDPR.

	EMAIL	LINK	PHONE	DATE	NATION	LOC	ORG	EIN	FAM	PNOUN	GNOUN	TOA	PER	HOB
Art. 4	✔	✔	✔	✔	✔	✔	✔	✔					✔	✔
Art. 9					✔				✔	✔	✔	✔	✔	

collection and processing of certain PSI types [37]. If no specific GDPR exception applies, processing PSI without the user's consent is prohibited (see article 9).

Article 9 is of significant importance for CV analysis as it prohibits processing personal data revealing sensitive information such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic and biometric data used for unique identification, health information, or data concerning a person's sex life or sexual orientation. These sensitive PSI types contribute little to assessing applicant expertise, making them a key target for de-identification. Article 4 of the EU GDPR defines personal data as any information that can be used to directly or indirectly identify a natural person, e.g., ID numbers, location data, online identifiers, or characteristics specific to their physical, mental, genetic, economic, cultural, or social identity. Table 2 maps our PSI types to GDPR art. 4 and 9 to show the relevance for GDPR-compliance.

Identifying suitable PSI detection techniques: We first discuss which different approaches are suitable to achieve the desired de-identification. Second, we assess the complexity of each PSI type and evaluate the suitability of various approaches, including regular expressions, AI-based methods, or dictionary-based techniques. The results are presented in table 3. Against our initial considerations, we have opted not to use anonymization, and instead we recommend the application of pseudonymization based on hash values where appropriate. Since CVs are typically not very long, the risk of identifying an individual from a pseudonym is minimized, while on the other hand, the information that the same entity is meant can be highly valuable in terms of data usability.

Email addresses (EMAIL) should be removed since they are direct identifier. Due to their static structure they are easy to detect with a regular expression.

Links (LINK) should be removed because it is a direct identifier, often linking to a networking platform, e.g. LinkedIn. We indicate a pseudonym option because this could improve readability if de-identification is guaranteed. URLs exhibit a static structure and are in general easy to detect with a regular expression.

Phone numbers (PHONE) of the CV owner should be removed because they are direct identifiers. Since we have decided that the previous employers should not be the reason for hiring, also phone numbers of previous employers should be removed. Rare cases might exist where recommendations by previous employers or inconsistencies in letters of recommendation need clarification. In this rare cases, a conflict management should be used to subsequently reveal numbers of interest. Using pseudonyms improves readability if de-identification is guaranteed.

Date (DATE) we generally promote to remove, to prevent discrimination by age Rigotti and Fosch-Villaronga [34]. European institutions like the European

Table 3. Evaluation of PSI types and suggested de-identification techniques from ISO/IEC 20889:2018-11 [17], ● marks our implementation recommendation.

TAG	Remove	Generalize	Pseudonymize	AI	Regex	Dictionary
EMAIL	●		●		●	
LINK	●		●		●	
PHONE	●		●		●	
DATE	●	●	●		●	
NATION	●			●		●
LOC		●	●	●		
ORG		●	●	●		
EIN		●	●	●		
FAM	●			●		
PNOUN		●				●
GNOUN		●				●
TOA	●					●
PER	●		●	●		
HOB	●			●		

Central Bank often use field experience, but it should be reasonably specified. For example, using date ranges such as 1–3 years, 3–10 years, or minimum experience such as more than 5 years of experience in software development offers less granular and more equitable criteria. Excessive demands, such as requiring 30 years of experience, are unreasonable and discriminatory. Because dates exhibit a static structure, they can be identified via regular expressions. Nevertheless, many date formats exists, thus it is important to include all of them. Pseudonomization can help to keep track of the same time period.

Nationalities (NATION) and all attributes towards race should be removed. According to [22, 7] the so-called "resume whitening" has demonstrated a considerable reduction on discrimination of minority groups. In general, nationalities and race can often be identified directly using dictionaries. However, this straightforward approach may overlook other indirect identifiers that could reveal nationalities or race, such as former employers listed in the CV and the CV owner. Therefore, more sophisticated methods, such as AI-based approaches, should be considered. Due methods' complexity, we have set nationalities out of scope.

Location details (LOC) can disclose significant information about the CV owner or their former employers. To still keep utility of data, we advocate generalization or pseudonomyzation. Especially generalization should be considered carefully as nationality could still be inferred, causing ethic bias. Given the diversity of location names, AI-based methods are most effective for this task.

Organizations (ORG) and *education institutes* (EIN) should be generalized or pseudonymized to prevent reputation-based hiring and unlink them from PSI types like nationality. As both represent institutional affiliations, merging them into a single "ORG" category allows for uniform pseudonymization, simplifying detection and processing while preserving linkability (e.g., repeated affiliations). Although conceptually distinct, their shared role as institutional entities justifies this approach. Hash-based pseudonymization prevents identification and reputation bias, but AI detection remains necessary due to contextual dependencies.

Detecting *relationship status* (FAM) in free text fields is challenging. Although specific lexical cues exist, automatic removal must consider the full context to be effective. In our further analysis, family status and relationships are not considered. However, it might be obfuscated, by removing family member names.

Following the approach of Cao and Daumé III [4] *pronouns* (PNOUN) should be generalized, e.g., "she" is replaced by "they" and *gendered nouns* (GNOUN) are generalized, e.g., stewardess is replaced by flight attendant. This can be implemented by using a dictionary. Due to the poor performance in generating gender related nouns, we excluded them from further investigation as explained below. Moreover, *terms of address* (TOA) should be removed [4] . This can be implemented using a dictionary pointing to an empty string. In general, it would also be possible to generalize Mr/Ms to Mx.

To remove *names* (PER), an AI-based approach is necessary due to the variety of names. A common issue are names in organization names, e.g., Mercedes. In such cases, an AI-based approach is required for context-based classification. The use of pseudonyms can be helpful to keep track of persons mentioned repetitively.

Hobbies (HOB) have no influence on job experience and expertise and should therefore be removed. We assume hobbies to appear in a respective section that is already labelled as hobbies. Thus, we will not further consider hobbies.

4.2 Demonstration

After several rounds of improvements, we derived a suitable prompt using GPT-4. For simplicity, we selected the following free-text fields, which are based on the Europass template but not identical: *Personal Statement, Work Experience, Education and Training, Personal Skills, Additional Information.* We manually checked each annotated CV for incorrectly assigned PSI. Unfortunately, we had to remove the gendered nouns because GPT-4 did not understand the task, and it generated for GNOUN, words such as "spacecraft design" or "aerodynamics" which are clearly not gender related. We also encountered that due to performance limitations, we can generate CVs only one by one. Thus, except for GNOUN, the synthetic generation was successful. The generated data was used to test the model performance in de-identifying PSI. We used the following prompt:

Objective: *Generate two synthetic CVs in JSON format, each belonging to individuals with backgrounds in technical and consulting fields, to be used as high-quality training data for a Named Entity Recognition (NER) task.*

Specifications:

1. Template Use: *Draw from the EUROPASS CV template, specifically utilizing sections where free text entry is permitted. These sections include: Personal Statement, Work Experience (with Descriptions), Education and Training (with Descriptions), Personal Skills, Additional Information*

2. Content Requirements: *Each CV must contain multiple descriptions within the free text fields, ensuring no section is left unfilled. Incorporate the following Personal Sensitive Information (PSI) across the text fields, using the stated abbreviations for annotation purposes: PER (Names), DATE (Dates), PNOUN (Pronouns), GNOUN (Gendered Nouns), TOA (Terms of Address), EMAIL (Email Addresses), LINK (Web Links), PHONE (Phone Numbers), LOC (Locations), ORG (Organizations). Ensure that the abbreviations are used within the entity fields of the annotations, not within the actual text.*

Table 4. Results of the PSI detection prototype tested on GPT-4 generated data.

	EMAIL	LINK	PHONE	DATE	PER	ORG	LOC	PNOUN	TOA
#	2	6	2	120	28	92	30	56	17
F1-score	1	1	1	0.89	1	0.73	1	0.56	1
Precision	1	1	1	1	1	0.57	1	0.39	1
Recall	1	1	1	0.8	1	1	1	1	1

3. **Realism:** *The synthetic data should closely mimic real-world CV data, avoiding generic placeholders (e.g., "John Doe").*
4. **Output Files:** *Annotated JSON (annotated_CV_data.json): Contains the text with entities annotated according to the specified structure. Plain JSON (CV_data.json): Includes only the text of the CVs without annotations.*
5. **Annotation Structure:** *For the annotated JSON, use the following format for each CV's sections, ensuring entities are correctly identified with start and end positions, types, and text: json Copy code { "CV1": { "PersonalStatement": [{ "text": "<text>", "entities": [{"start": <start>, "end": <end>, "type": "<entity type>", "text": "<entity text>"}] }] } } Plain Data Structure: For the plain JSON, structure it as follows, excluding entity annotations: json Copy code { "CV1": { "PersonalStatement": [{"text": "<text>"}] } }*
6. **Delivery:** *Generate JSON files in a human-readable manner and provide them in json format in the chat. There should be no need for displaying additional information or instructions in the chat beyond this.*

Prototype: We provide insights into the results of our algorithm. Following ISO/IEC 20889 [17] we use hashing as our standard pseudonymization technique. This allows us to assign the same hash to e.g., organizations mentioned several times in the document. Following table 3 we also implemented regular expressions and dictionaries. Model architecture and testing process are provided in Figure 1.

As evaluation metrics, we use precision, recall and f1-score. Especially recall is of high importance when de-identifying PSI due to the mandatory requirement of ensuring that all PSI is effectively removed or anonymized. Thus, in the CV de-identification scenario, recall is defined as the ratio of actual PSI that has been successfully identified and de-identified by our algorithm. The evaluation results of our detection approach are provided in Table 4. In the following, we will present some examples of our de-identification approach (see table 5), structured by PSI types from table 1. We have chosen DATE, PNOUN, TOA, PER and ORG since they represent each PSI detection technique. Each example shows the original input alongside its anonymized version, illustrating how PSI is replaced or generalized to ensure privacy while preserving the contextual structure.

Following our de-identification strategy from Figure 1, we have defined several regex definitions to cover multiple *date* (DATE) types. Table 5 shows a date span that is separated with a "to" instead of a "-". Especially if further words are added between the date span e.g. "to the beginning of" a simple regex function might fail to combine the dates into one tag. While for de-identification this might not be a big deal, it might influence the accuracy metrics if the NER results differ from the actual data.

As part of our gender de-identification approach we generalize *pronouns* (PNOUN) into a neutral form as highlighted (pink) in Table 5. We implemented a conservative approach utilizing a dictionary of pronouns. This approach only

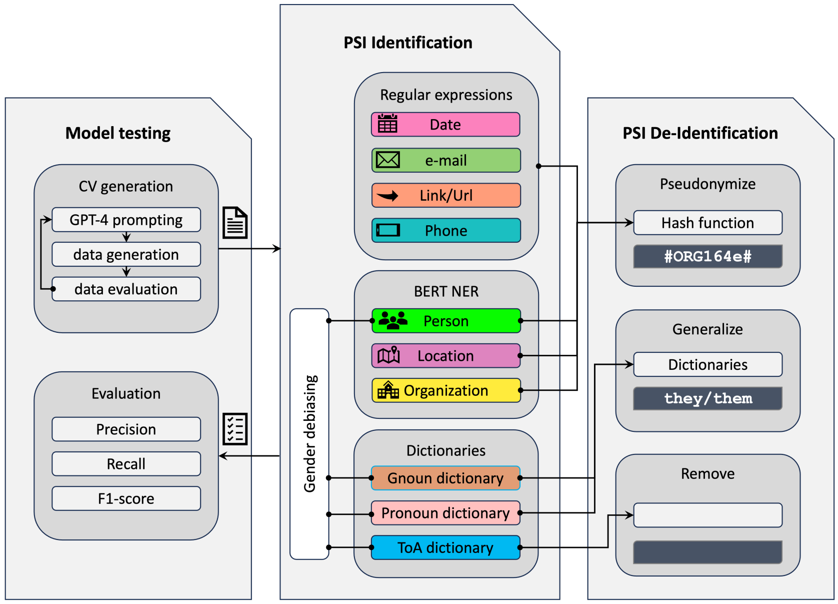

Fig. 1. Model testing, PSI identification and PSI De-identification approach.

Table 5. Selected de-identification examples of ORG, DATE, TOA, PNOUN, PER.

PSI	Original	Modified
ORG DATE	Consulting Analyst at BizMax Solutions from April 2014 to February 2019 .	Consulting Analyst at #ORG164e# from #DATE8ce9# to #DATE7d15# .
TOA PNOUN	Accomplished financial consultant (Mr.) with extensive experience in financial restructuring, risk management, and strategic investment planning. Known for his (he/him) ability to deliver financial solutions.	Accomplished financial consultant with extensive experience in financial restructuring, risk management, and strategic investment planning. Known for their (they/them) ability to deliver financial solutions.
PER TOA	Data Scientist Alice Johnson (Ms.) specializing in statistical analysis and predictive modeling.	Data Scientist #PER4fa8# specializing in statistical analysis and predictive modeling.

covers common pronouns and fails if neopronouns are used, especially if they are fantasy based words, e.g. vix/vixelf. Training an AI on such neopronouns is a promising solution to be further researched. We remove *terms of address* (TOA) as highlighted (cyan) (see Table 5). We again use a dictionary that contains all relevant TOA. Unexpected characters around the term of address should be removed to improve readability and PSI detection accuracy.

To de-identify *names* (PER), we utilize the bert-large-NER model and its PER tag (see Table 5). As mentioned in the organization section, issues in detection can arise if a natural person's name is used as an organization name.

Table 6. Retained Utility and Readability of Tags, High (H), Medium (M), Low (L).

Category	Retained Utility			Readability		
	H	M	L	H	M	L
Personal Statement	1	0	0	0.55	0.45	0
Work Experience	0.85	0.1	0.05	0.05	0.85	0.1
Education and Training	0.45	0.25	0.3	0.1	0.8	0.1
Personal Skills	1	0	0	0.95	0.05	0
Additional Information	0.9	0.1	0	1	0	0

To de-identify *organizations* (ORG), we use the pre-trained BERT model bert-large-NER from Huggingface (2023). A common issue is the correct separation of organizations that consist of several nouns. Especially if names are part of the organization, this often resolved as a name tag. Uncommon organization names, e.g. "DevCon", are missclassified as MISC where BERT collects all entities, not labeled as ORG, PER or LOC. Thus, when training a BERT model, PSI types event and education should be considered.

Regarding retained usability we achieved a cohens kappa of 0.64 (substential agreement) and for readability a cohens kappa of 0.46 (moderate agreement). A reason for the reduced kappa in readability might be a different perception of how the reader is interrupted. Overall, we find a reduced readability and utility for *Work Experience*, and *Education and Traing*. This can be explained by excessive annonymization, de-identifying important parts, e.g. job roles (see Table 6).

5 Discussion

In this work, we aim to motivate companies and job posting platforms to implement CV de-identification by reducing implementation hurdles. Additionally, our approach supports companies and job platforms in achieving GDPR and AI Act compliance. We provide a starting point to answer which and how data should be de-identified. This groundwork is vital for future AI applications, as AI trained on de-biased, high-utility data is promising to operate with reduced bias.

5.1 Implications for Industry

Both, bias in human decisions caused by prejudice and bias trained from human behavior are both unacceptable and shold both be mitigated. Especially for model training, removing bias in advance while keeping data utility is one of our key objectives, to achieve a more effective alignment between applicants and job roles [7]. In the following, we communicate strategies and considerations:

Bias and unfairness are especially caused by PSI that are mentioned in article 9, GDPR and should not be considered for CV classification tasks. Models can profit from a better matchmaking if those PSI are de-identified.

Linkability issues are often caused by PSI from article 4, GDPR. Here, de-identification can reduce the risk of non-compliance and for a fair CV assessment.

Utility describes the usefullness of data. Pseudonymization and generalization, if applied properly, can help to maintain utility while at the same time de-identifying PSI. If applied, risks and benefits must always be carefully assessed. It should always be aimed to achieve full functionality [5] while simultaneously de-identifying data effectively.

Removal should be considered in case of uncertainties or extensive occurrence of the same PSI to break linkability or bias. This can cause reduced utility.

Pseudonymization enables linking of associated records within a CV [17]. If applied, the trade-off between de-identification and utility must be considered. Hash values for de-identification can be crucial to correctly de-identify a persons first and last name and link them to exactly one hash value. Correct start and stop values of the PSI are mandatory.

Generalization reduces granularity and allows preserving data truthfulness at the record level [17]. If used, it is important to consider the trade-off between de-identification and utility. Thus, choosing the right granularity is important.

Neopronouns are difficult to detect and can cause applicant identification, especially if rarely used.

Readability of de-identified CVs is important for human processing. Removing data completely may break the structure of a sentence. However, too many tokens in the data can decrease readability. Thus, it might be useful to summarize consecutive tokens for human readability.

5.2 Regulatory implications of the EU AI ACT

The EU Artificial Intelligence Act [2], one of the first regulations of its kind, classifies AI systems into four risk levels: *unacceptable* (prohibited, e.g., social scoring), *high-risk* (heavily regulated, e.g., healthcare AI), *limited risk* (subject to transparency obligations, e.g., chatbots), and *minimal risk* (unregulated, e.g., AI-enabled video games). AHS, including AI-driven CV classification tools, are likely to be classified as high-risk due to their significant influence on employment decisions and their potential impact on individuals' rights. Consequently, these systems must meet strict regulatory requirements before deployment in the EU market, ensuring transparency, fairness, and bias mitigation. Ensuring compliance with the AI Act requires addressing two key aspects. First, AHS tools must be trained on high-quality, unbiased datasets to prevent unfair hiring practices. In this regard, CV de-identification plays a crucial role by ensuring that AI models assess candidates solely on skills and experience rather than demographic attributes, effectively reducing bias and enhancing fairness. Second, automated hiring decisions cannot be entirely autonomous, human oversight is mandated to review and override AI recommendations when necessary. However, human intervention does not eliminate bias either. Particularly during the initial screening phase, CV de-identification helps mitigate unconscious bias, ensuring that hiring decisions remain fair, objective, and merit-based. By prioritizing fairness and transparency, CV de-identification aligns automated hiring systems with the EU AI Act's regulatory framework, mitigating compliance risks while fostering equitable recruitment practices.

5.3 Socio-technical framing

The adoption of CV de-identification tools depends on organizational culture, recruiter practices, and existing hiring processes. Recruiters may resist using de-identified CVs if they perceive personal information as necessary for assessing cultural fit or networking potential. Thus, integrating such tools requires raising recruiters' awareness of fairness and bias issues while balancing their information needs, making organizational training and change management essential for successful adoption. However, although de-identification enhances fairness and compliance, it may reduce recruiters' confidence or perceived completeness of a candidate's profile. Providing clear guidelines that explain the rationale of de-identified CVs, and embedding de-identification into existing hiring workflows is essential to promote acceptance and effective use.

5.4 Limitations and future work

While our research contributes valuable insights into de-identifying CVs to enhance fairness during the job screening phase of recruitment, subsequent recruitment phases, including interviews and final selection processes, fall outside the scope of this paper. Thus, potential biases that may arise during in-person interactions or decision-making stages not covered here, serve as opportunities for future work.

Given the current limitations of GPTs in swiftly generating synthetic testing data, our approach necessitates training our own AI model that can recognize fine-tuned CV specific PSI types such as education institutes or nationalities. We still evaluate traditional annotation approaches as more time-consuming since tremendous amounts of time and several annotators following well-structured annotation guidelines are required [1] Although we have elicited the most common PSI types from the GDPR and existing literature, some rare PSI types with high risk data might exist that are not considered yet, e.g. medical diagnosis. Whether an evaluation of all existing PSI types can be implemented depends on costs of implementation, maintenance or accuracy. Similar to Jensen et al. [19] who used real world data, some PSI types are not well represented to provide comprehensive testing results. This limitation could be solved by using larger data sets or oversampling techniques. As a next step, we aim to addres the above-mentioned limitations by implementing an advanced AI based de-identification prototype, based on fine-tuned NER trained, on synthetic training data.

6 Conclusion

In this paper, we present valuable insights on how to de-identify PSI in CVs. We elicited relevant PSI types that need to be considered in the de-identification of CVs. We provide valuable insights into the appropriate de-identification approach for different PSI types. A first PSI de-identification prototype is implemented using synthetic data, generated using GPT-4. With this research, we aim to motivate companies and job platforms by reducing the implementation hurdles of CV de-identification. In doing so, this work contributes to more fairness and de-biasing in the job market.

Bibliography

1. Al Aziz, M.M., Ahmed, T., et al.: Differentially private medical texts generation using generative neural networks. ACM Transactions on Computing for Healthcare 3(1), 1–27 (2021)
2. Artificial Intelligence Act: Proposal for a regulation of the european parliament and the council laying down harmonised rules on artificial intelligence and amending certain union legislative acts. EUR-Lex (2024)
3. Baumann, J., Castelnovo, A., Cosentini, A., Crupi, R., Inverardi, N., Regoli, D.: Bias on demand: Investigating bias with a synthetic data generator. In: Elkind, E. (ed.) Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23. pp. 7110–7114 (8 2023)
4. Cao, Y.T., Daumé III, H.: Toward gender-inclusive coreference resolution. arXiv preprint arXiv:1910.13913 (2019)
5. Cavoukian, A.: Privacy by design (2009)
6. Chua, H.N., Ooi, J.S., Herbland, A.: The effects of different personal data categories on information privacy concern and disclosure. Computers & Security 110, 102453 (2021)
7. Deshpande, K.V., Pan, S., Foulds, J.R.: Mitigating demographic bias in ai-based resume filtering. In: 28th ACM conference on user modeling, adaptation and personalization. pp. 268–275 (2020)
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
9. Dresch, A., Lacerda, D.P., Antunes, J.A.V.: Design Science Research, pp. 67–102. Springer International Publishing, Cham (2015)
10. Europass: Europass cv editor, european union (2024), https://europa.eu/europass/, accessed 2024-03-24
11. Frauendorfer, D., Schmid Mast, M., Sutter, C.: To include, or not to include? accuracy of personality judgments from resumes with and without photographs. Swiss Journal of Psychology 74(4), 207 (2015)
12. Giomi, M., Boenisch, F., Wehmeyer, C., Tasnádi, B.: A unified framework for quantifying privacy risk in synthetic data. Proceedings on Privacy Enhancing Technologies 2, 312–328 (2023)
13. Gu, J., McFerran, B., Aquino, K., Kim, T.G.: What makes affirmative action-based hiring decisions seem (un) fair? a test of an ideological explanation for fairness judgments. Journal of Organizational Behavior 35(5) (2014)
14. Gualdi, F., Cordella, A.: Theorizing the regulation of generative AI: lessons learned from italy's ban on chatgpt. In: 57th HICSS 2024, (2024)
15. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS quarterly pp. 75–105 (2004)
16. Hu, Y., Wu, F., Li, Q., Long, Y., Garrido, G.M., Ge, C., Ding, B., Forsyth, D., Li, B., Song, D.: Sok: Privacy-preserving data synthesis. In: 2024 IEEE Symposium on Security and Privacy (SP). pp. 4696–4713 (2024)

17. ISO/IEC 20889:2018-11: Privacy enhancing data de-identification terminology and classification of techniques. Standard, International Organization for Standardization, Geneva, CH (Nov 2018)
18. Javed, F., Luo, Q., McNair, M., et al.: Carotene: A job title classification system for the online recruitment domain. In: International Conference on Big Data Computing Service and Applications (2015)
19. Jensen, K.N., Zhang, M., Plank, B.: De-identification of privacy-related entities in job postings. arXiv preprint arXiv:2105.11223 (2021)
20. Ji, Z., Lin, C., Wang, H., Shen, C.: Speech-forensics: Towards comprehensive synthetic speech dataset establishment and analysis. In: Larson, K. (ed.) Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. pp. 413–421 (8 2024), main Track
21. Jiang, X., Zhou, X., Grossklags, J.: Privacy-preserving high-dimensional data collection with federated generative autoencoder. Proceedings on Privacy Enhancing Technologies 2022(1), 481–500 (2022)
22. Kang, S.K., DeCelles, K.A., Tilcsik, A., Jun, S.: Whitened résumés: Race and self-presentation in the labor market. Administrative science quarterly (2016)
23. Karl, F., Scherp, A.: Transformers are short-text classifiers. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction. pp. 103–122. Springer (2023)
24. Kowalczyk, P., Röder, M., Rottmann, J., Thiesse, F.: Designing a method to nudge analytics with artificially generated data. In: Proceedings of the International Conference on Information Systems (ICIS) (2023)
25. Löbner, S., Tesfay, W.B., Bracamonte, V., Nakamura, T.: Systematizing the state of knowledge in detecting privacy sensitive information in unstructured texts using machine learning. In: International Conference on Privacy, Security and Trust (PST) (2023)
26. Löbner, S., Tronnier, F., Linke, D.: De-identification of privacy sensitive information in resumes with gpt-4: An utility analysis for automated job role classification. In: Proceedings of the Hawaii International Conference on System Sciences (2025)
27. Memmert, L., Cvetkovic, I., Bittner, E.: Human-ai collaboration in conceptualizing design science research studies: Perceived helpfulness of generative language model's suggestions. In: ECIS (2023)
28. Pal, R., Shaikh, S., Satpute, S., Bhagwat, S.: Resume classification using various machine learning algorithms. In: ITM Web of Conferences (2022)
29. Pape, S., Serna-Olvera, J., Tesfay, W.B.: Why open data may threaten your privacy. In: Workshop on Privacy and Inference. pp. 1–5 (2015)
30. Parasurama, P., Sedoc, J.: Degendering resumes for fair algorithmic resume screening. arXiv preprint arXiv:2112.08910 (2021)
31. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of management information systems (2007)
32. Raghavan, M., Barocas, S., Kleinberg, J., Levy, K.: Mitigating bias in algorithmic hiring: Evaluating claims and practices. In: 2020 Conference on Fairness, Accountability, and Transparency (2020)

33. Ribeiro, B., Rolla, V., Santos, R.: Incognitus: A toolbox for automated clinical notes anonymization. In: Conference of the European Chapter of the Association for Computational Linguistics (2023)
34. Rigotti, C., Fosch-Villaronga, E.: Fairness, ai & recruitment. Computer Law & Security Review (2024)
35. Schneider, S., Steuber, F., Schneider, J.A.G., Rodosek, G.D.: How good can machine generated texts be identified and can language models be trained to avoid identification? In: HICSS (2024)
36. Skondras, P., Psaroudakis, G., Zervas, P., Tzimas, G.: Efficient resume classification through rapid dataset creation using chatgpt. In: 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA). pp. 1–5. IEEE (2023)
37. Tesfay, W.B., Serna, J., Rannenberg, K.: Privacybot: detecting privacy sensitive information in unstructured texts. In: SNAMS. IEEE (2019)
38. Tesfay, W.B., Serna, J.M., Pape, S.: Challenges in detecting privacy revealing information in unstructured text. In: PrivOn@ ISWC (2016)
39. Tronnier, F., Löbner, S., Lacombe, M.H., Rannenberg, K.: Regulatory challenges in cybersecurity – a critical analysis of the eu ai act. In: Information Security Education. Empowering People through Information Security Education. Springer Nature Switzerland (May 2025)
40. Tronnier, F., Pape, S., Löbner, S., Rannenberg, K.: A discussion on ethical cybersecurity issues in digital service chains. In: Cybersecurity of digital service chains: challenges, methodologies, and tools. Springer International Publishing Cham (2022)
41. Tronnier, F., Löbner, S., Azanbayev, A., Walter, M.L.: A systematic literature review on gender bias in ai – towards inclusiveness in machine learning. In: Proceedings of the Pacific-Asia Conference on Information Systems (2024)
42. Umbach, R., Henry, N., Beard, G.F., Berryessa, C.M.: Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. In: Proceedings of CHI 2024. CHI '24, New York, NY, USA (2024)
43. Zaniboni, S., Kmicinska, M., Truxillo, D.M., et al.: Will you still hire me when i am over 50? the effects of implicit and explicit age stereotyping on resume evaluations. European journal of work and organizational psychology 28(4), 453–467 (2019)
44. Zaroor, A., Maree, M., Sabha, M.: Jrc: a job post and resume classification system for online recruitment. In: 29th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE (2017)
45. Zhao, W., Song, S., Zhou, C.: Generate synthetic text approximating the private distribution with differential privacy. In: Larson, K. (ed.) Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24. pp. 6651–6659 (8 2024)
46. Zhou, J., Zhang, Y., Luo, Q., Parker, A.G., De Choudhury, M.: Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23, New York, NY, USA (2023)