Data Donation: A First Step Towards Improving Representativeness in Algorithmic Hiring Datasets

Jorge Saldivar^a*, Jessica Wulf^b, and Carlos Castillo^a

^aSchool of Engineering, Universitat Pompeu Fabra, Barcelona, Spain;

^bCenter for Responsible Research and Innovation, Fraunhofer Institute for Industrial Engineering IAO, Berlin, Germany

*corresponding author: jorge.saldivar@upf.edu

Data Donation: A First Step Towards Improving Representativeness in Algorithmic Hiring Datasets

The use of algorithmic hiring is on the rise, becoming more and more common due to its capacity to process a large number of applications with efficiency. Despite the advantages of this approach, there have been numerous cases of discriminatory hiring outcomes. A significant contributing factor to such outcomes is the lack of representativeness in the data used to develop these systems. This results in a significant decline in performance for underrepresented groups, disproportionately impacting marginalized communities.

Addressing bias in algorithmic hiring requires access to comprehensive datasets that include curriculum vitae (CV) and demographic information reflecting diverse backgrounds. Unfortunately, there is a lack of datasets that serve such a purpose. This paper introduces a data donation campaign designed to collect real-world CVs, including demographic and sensitive information, from a representative sample of individuals. The campaign aims to establish a foundational dataset for the generation of synthetic CVs that enable the development of fairer and more inclusive algorithmic hiring tools.

The paper discusses the design decisions underpinning the campaign, along with the challenges encountered during its deployment and execution. Finally, it offers lessons learned and practical solutions to overcome these challenges, thereby contributing valuable insights for future efforts in this domain.

Keywords: Data donation; dataset representativeness; algorithmic hiring; discrimination

Introduction

Algorithmic hiring (i.e., the use of Artificial Intelligence (AI) in recruitment) is on the rise with particular prevalence in some sectors.; job postings that used to attract about 120 applicants in 2010 now attract over 250 (Fuller et al., 2021). AI technologies promise to deal with hundreds or thousands of applicants at high speeds. Moreover, their uptake in European HR teams and Public Employment Services is growing faster than the global average (European Commission, 2019).

Despite its promises, however, discriminatory outcomes have been documented in almost every domain of AI application, including algorithmic hiring (Köchling and Wehner, 2020). A critical reason for discrimination in AI-based tools is the lack of representativeness in the training data (Barocas & Selbst, 2016), as the system performance drastically degrades for underrepresented collectives, negatively impacting marginalized groups (Portela et al., 2024).

The research on discovering, measuring, and mitigating bias in hiring processes assisted by AI algorithms requires using real-world curriculum vitae (CV) and demographic data that reflects the characteristics of people from diverse demographic backgrounds. Nevertheless, there is a lack of datasets on these characteristics, which can be attributable to the fact that CVs typically contain sensitive information. The available data is, however, composed of synthetic CVs and similar materials fabricated using aggregated, incomplete, or artificially annotated demographic data (Fabris et al. 2024). Moreover, the collection of sensitive information, including sexual orientation, religion or belief, and ethnicity, is a challenging process, in part due to legal constraints that limit the access, storage, and use of such data (Finck, 2020).

To address the lack of representativeness in algorithmic hiring datasets, it was determined that a digital data donation approach could be a valuable tool. Digital data

donation is the process by which individuals provide their personal data via technology to a third party for use in specific contexts, typically for academic research purposes. This practice has been extensively used in the healthcare domain but has also been applied in other fields, such as politics and computer science (Puschmann, 2019).

We describe in detail a data donation campaign that collected over 1,000 real-world CVs and the corresponding demographic and sensitive data of the CV's owners. The collected data are intended to serve as a reference for the generation of synthetic CVs, aiming to create as much representative dataset as possible for the EU workforce. This envisioned dataset of synthetic CVs will facilitate the development of fairer algorithmic hiring tools. This paper provides a comprehensive discussion of the challenges encountered during the design phase of the data donation campaign, as well as a detailed account of the implementation process, including the instruments and procedures employed.

The paper is structured as follows: Section 2 presents background concepts while Section 3 introduces the relevant literature. Section 4 describes the method. In Section 5, we explain the results, while Section 6 discusses the challenges and lessons learned, and Section 7 presents concluding words.

Background

The donation of data for research is a long-lasting practice in health scholarship, and its value is highly recognized for addressing questions about diagnosis, prevention, and therapies for unusual and chronic diseases (Weitzman et al., 2010). Yet, its application to enable and advance research in fields like social and computer science has lately gained attention (Strotbaum *et al.*, 2019; Couto, Garimella, 2024).

An increasing number of studies ranging from discussions about the ethical, legal, and organizational aspects of data donation (Strotbaum *et al.*, 2019; Couto and Garimella, 2024) to proposals on methods, frameworks, and tools to effectively collect donations (Loecherbach,

2022), to investigations into the motives that drive people to engage into data donation initiatives (Skatova and Goulding, 2019) are currently available in the literature. Next, state-of-the-art studies on data donation are reviewed and discussed.

State of the art

A large body of research on data donation has focused on exploring the challenges of designing, organizing, and conducting personal data collection initiatives. A common challenge identified by Gomez Ortega et al. (2021) and Hummel et al. (2019), as well as by Weitzman et al. (2010) and Keusch et al. (2024), is the necessity to build trust between donors and the recipient institution or research team. Having trust in the donation recipient has been identified as an essential driver in increasing the willingness of people to donate their data (Skatova and Goulding, 2019). In this sense, Garimella and Chauchard discuss strategies for approaching donors to foster trust (i.e., face-to-face vs. online) (2024). In turn, Kuiper and Akdag propose to maximize trust by improving the user interface of tools used in data collection (2021). Similarly, Maus et al. introduce a prototype mobile application that aims to enhance trust by properly communicating information about the recipient, privacy protection mechanisms, and data usage (Maus et al. 2020).

Donors' privacy and transparency in data usage and access are other challenges also discussed in the literature (Patterns, 2022). In this sense, questions explored by scholars include: How will data be pseudo- or anonymized? Are the existing technical and legal mechanisms sufficient to completely protect donors' identities? How data will be used, who can access it, and which qualification or training is required to work with donations (Strotbaum et al., 2019)? Along these lines, the representativeness of donated data has been a concern addressed by Bietz and colleagues (2019). Connected with these questions, Skatova et al. have explored the general public's attitudes toward data donation for research (2019). They found that researchers should emphasize the value and potential impact of the data in the study. Also, the authors discovered that donors are required to have control over their data, to be able to grant permissions over the data at the maximum possible granularity, and to use it in their prospective usage.

Various authors have proposed frameworks and tools for collecting data in the context of donation campaigns. In this sense, Boeschoten and colleagues introduce a procedure to facilitate the collection of "data download packages (DDP)," i.e., personal digital trace data packages that modern platforms, like social media or search engines, ought to provide to users. Their workflow guides donors in downloading their digital traces from these platforms (e.g., Instagram, Facebook, YouTube), minimizing the DDP to only relevant data and providing consent on the data (Boeschoten et al., 2020). A proof-of-concept of Boeschoten's workflow has been implemented in a mobile application and presented in (2022). Similarly, Araujo et al. worked on an open-source web-based framework under the design principles of i) data minimization, restricting access to the data that is strictly necessary for the study; ii) transparency, allowing donors to visualize specifically which data out of the entire DDP is going to be donated giving them the possibility to remove items from donation); and iii) flexibility, enabling customization on the configuration and operation of the framework depending on the research requirements (Araujo et al. 2017). A tool to support the donation of DDP from WhatsApp, easing the research on this platform, has been implemented via the web application WhatsApp Explorer and introduced by Garimella and Chauchard (2024). Besides DDP, data donation is also enabled through software installed on the donor's device (e.g., mobile phone, computer, wearables). Although more intrusive than DDP-based approaches, these applications to register user activities have also been proposed as a method to donate data, especially in health and behavioral research (Ohme et al., 2020; Araujo et al., 2017; Christner et al., 2022). An interesting discussion on the pros and cons of the different methods proposed in the literature to conduct data donation is presented by Ohme et al. (2024; van

Driel et al., 2022), who highlight the technical difficulties imposed on donors by each approach.

The application of data donation has started to be studied in other fields besides health research. For example, Puschmann (2019), as well as Couto and Garimella (2024) have reported using data donation to understand the consumption and access to political content on digital platforms and social media. Nevertheless, no studies to date have examined the potential of data donations to develop more representative data sets, for example, for algorithmic hiring.

Drawing on these efforts, this work contributes to the state of the art by enriching the discussion about data donation as a method to obtain essential data for research on pressing topics that are currently impacting society, like representative datasets for algorithmic hiring. In what follows, we present our data donation approach exploring how the identified challenges (e.g., trust, transparency, privacy, technical issues, representativeness) have been addressed.

Material and Method

This section outlines the step-by-step approach to the data donation, including target group, materials (website and online survey), donation procedure, and design decisions. Furthermore, the identified challenges of data donations, like donor trust, transparency about data handling, or privacy concerns related to sensitive data (as discussed previously), are also addressed.

Target Group

We invited residents of the European Economic Area and Switzerland who are part of the labor force (i.e., employed or seeking employment) to voluntarily donate their CVs by completing an online survey. To use the CVs donated as a reference dataset, we set the goal of collecting at least 1000 CVs from individuals with diverse social backgrounds, professional sectors, and levels of seniority.

Website

We integrated the data donation information and survey into the official website of the research project (project name withheld for anonymous review) to increase transparency for donors and enable them to easily find more information on the project and the organizations that are part of it. On a section of the website, we summarized the topic, included a link to FAQs about the data donation, and stated that in the next step, the donor would be asked to upload an anonymized CV and answer optional questions about sensitive data. Furthermore, we integrated a link to the detailed privacy information and a fold-out block containing the information sheet (see Appendix 1).

The FAQs were a central part of the campaign website—in every step, the donor could open the FAQs in case of questions or insecurities. The FAQs included questions about processing of personal data (*What information do I give to you when donating my CV? What happens to my CV when submitting the online form?*), the project (*What is [project name]? Who is behind it?*), information on anonymity (*How can I anonymize my CV? Can someone find me through the data set?*), practical questions (*What information in my CV is relevant for you? In what formats can I upload my CV?*) and about the sensitive data questions (*Why am I asked about this? How did you develop the sensitive data questions?*).

Online Survey

Prior to the completion of the online survey, donors were required to confirm their understanding of the information sheet regarding the campaign, acknowledge their voluntary participation, and verify their fulfillment of the campaign's requirements. These requirements included being over the age of 18, residing in one of the countries of the European Economic Area or Switzerland, and being part of the labor force. In addition, the donors were required to consent to the utilization of the information contained within their CVs, as well as their demographic data. Subsequent to the completion of the requisite consent form, donors were presented with an online survey consisting of three sections: (i) the uploading of CV(s), (ii) the designation of job categories and professional seniority, and (iii) a questionnaire concerning sensitive data (see Appendix 2).

The first section of the online survey (see Appendix 2a) entailed the upload of one or more CVs. Donors were asked to anonymize their CVs by deleting all identifiable information about them, including their name, home address, phone number, email address, and/or personal picture. Donors were asked to submit a CV they would have no problem posting online. In instances where the CV was written in a second language, donors were invited to submit the translated version as well, following the removal of identifiable information.

To reach individuals with diverse backgrounds and work histories, we included the possibility of manually entering their educational background, professional experience, and skills in an open-text field. This approach was designed to include donors who might not have a CV due to its infrequent use in their professional domain or those who prefer not to upload their CVs for reasons of privacy.

Subsequently, the participants were requested to specify the professional field or fields for which the donated CVs are pertinent. The listed categories were taken from a combination of the ESCO first- and second-level categories (https://esco.ec.europa.eu/en). In this part of the survey, donors were invited to provide information about their seniority, age, and highest level of education achieved (see Appendix 2b).

The final section of the online survey comprised a questionnaire concerning demographic and sensitive data (see Appendix 2c). In the formulation of this questionnaire, considerable effort was invested in ensuring inclusivity and respect for individuals from diverse demographic backgrounds and social groups, particularly those who are marginalized. The objective was to create a questionnaire where respondents feel comfortable providing their responses. The development of this questionnaire was undertaken in accordance with the prevailing best practices for the collection of equality data as recommended by the scientific community, European Union Agency for Fundamental Rights, and European Network Against Racism (ENAR, 2016; Badgett, 2009; Palmer and Harley, 2012; Baumann, Egenberger and Supik, 2018). Accordingly, in light of these recommendations, the principle of voluntariness was ensured to be in effect: there was always the option *Prefer not to answer* available. Furthermore, all answering options allowed self-identification. As an example, for gender identity, the following two questions were asked: *What gender do you identify with?* (*options=[Prefer not to answer; Woman, Man, No-binary, I prefer to self-identify (open text field)]*) and *Does your gender differ from the gender assigned at birth? (options=[Prefer not to answer; Yes, I identify with a gender different from the one I was assigned at birth; No, I identify with the same gender I was assigned at birth; Other]*).

Furthermore, anti-discrimination experts recommend including questions about the subjective experience of discrimination, especially regarding experiences of racism (many questionnaires use questions about migration experiences as a proxy for racism, which is not recommendable) (Baumann et al. 2018). Accordingly, ethnicity was asked using the following two questions: *Are you perceived as `foreign" or not ``white" in the country where you live?* (options=[Yes, often; Yes, sometimes; No, never]) and Independently of what you answered to the previous question, would you self-describe as belonging to an ethnic minority in the country where you live? (options=[I consider myself within the ethnic majority in the country where I live; I consider myself as belonging to both (an) ethnic minority and (an) ethnic majority]).

The decision was made to inquire solely about the characteristics that have been identified in previous research as typically being inferred from CVs by recruiters and that may generate bias for or against certain candidates (EIGE, 2021; Europen Commission, 2020; Chander, 2017; Kovacheva *et al.*, 2018). In accordance with the preceding research, inquiries were made regarding the following sensitive information: age, gender, sexual orientation, race and ethnicity, religion, and disability status. The first draft of the questionnaire was shared with an expert on sensitive data and anti-discrimination questions, as well as the consortium partners working closely with marginalized groups. The feedback received was included in the final questionnaire.

The campaign forms were available in four languages: English, Spanish, Catalan, and German. The questionnaire regarding sensitive data was examined by individuals who are proficient in these languages as their primary language and possess an understanding of anti-discrimination discourse within this context.

Donation Procedure

The donation procedure began with the recruitment of donors online, who were subsequently directed to the designated donation landing page (see Appendix 1). Upon arriving at the designated donation website, potential donors were provided with an informational sheet and were instructed to complete the consent form. Following the acquisition of consent, donors advanced to the upload form. Following the removal of all identifiable information, the donors were asked to upload one or more CVs as a PDF or DOC/DOCX/ODT file or alternatively share their educational background, professional experience, and skills in open text fields.

Next, donors were asked about their work experience (i.e., job sector, seniority, highest level of education attained). Later, respondents were asked a series of optional questions regarding age, gender, and sensitive information like their sexual orientation,

religion, ethnicity, and disability condition. When finishing the questionnaire, participants submitted their donations. In response, donors received a message thanking them for participating and instructing them how to withdraw from the campaign at a later point. Since participation was anonymous by design—no personal information was collected through the survey—the submission timestamp is included in the withdrawal instructions, asking to provide it when communicating the intention to be excluded from the campaign.

As previously mentioned, the objective of the collected data is to serve as a reference for the generation of synthetic CVs. The generative procedure that will be employed to create the synthetic CVs is outlined in Appendix 3.

Ethical Approval

The Ethics Review Board (ERB) at (university name withheld for anonymous review) approved the data donation campaign. After two rounds of reviews and revisions, the data collection methods—including digital instruments, promotion strategies, target groups, questionnaires, and informed consents—the data protection impact assessment, data processing, scope, and management mechanisms, as well as the procedure to safeguard donors' identity and approach to generate the synthetic data, were accepted by ERB certifying that the proposed protocol complies with the international ethical principle in research and is adjusted to European and local regulations on personal data protection.

Outreach Campaign

Fostering transparency in the data donation process is critical to incentive participation (Ohme et al., 2023; Weitzman et al., 2010). In this sense, we strived to explain our intention with the campaign by clearly defining the problem and its relevance for the general public and describing the process step by step. We aimed to communicate clearly, using visualizations (see Figure 1) and simple language to explain issues that, as noted in the literature, can

discourage participation if not adequately addressed. These include measures to protect donors' identities, mechanisms for managing data, and the intended uses and scope of the donations. (Strotbaum *et al.*, 2019). As previously mentioned, a concise and accessible frequently asked questions (FAQs) section was developed and incorporated into the donation website, ensuring its availability throughout the donation process.



Figure 1. Graphic included in the FAQs section to explain the management, intended usage and scope of the donated data.

A series of initiatives were implemented to foster a sense of trust among the parties involved in the donation process and encourage participation (Ortega, Bourgeois and Kortuem, 2021; Weitzman et al., 2010). The donation was linked to the research project, as the campaign was run directly on the project's website. Alternatives to the project's website, such as using survey-specialized web applications, were analyzed. However, the advantages of running the data donation on the project's website were convincing: access to information about the institutions and people executing the donation campaign were facilitated, furthermore, information about the funding agencies, project goals, and expected results could be found easily for the donors.

Furthermore trust was established by informing prospective donors of the fact that the donation protocol has been accepted by an institutional review board belonging to one of the consortium's academic partners. This acceptance was granted subsequent to the successful

completion of a rigorous review process (described in "Ethical Approval"). Supporting documents and approval certificates resulting from the ethical review process were publicized with the campaign information.

Reaching diversity and representativeness in donated data are concerns raised by the scientific community when organizing and running data donation campaigns (Bietz, Patrick and Bloss, 2019). In this sense, we applied multiple promotion strategies to reach a diverse target group, starting by carefully designing social media campaigns for platforms like LinkedIn, Twitter, and Mastodon. The posts were written according to the platform's specifications: the LinkedIn post, for example, included a visual, and Mastodon posts included only a short text and a link to the data donation website. The short text started with a question to the reader: *"Are you worried that you could be discriminated against when looking for a job?"* The objective was to address the reader directly and demonstrate the daily relevance of the topic. Next, to support the reader, the answer to the question was also given: *"You're not alone!"* The rationale behind the response was to underscore the societal pertinence of the subject matter. The middle of the post included information about algorithmic hiring and discrimination: *"More and more companies are using recommendation systems based on #algorithms to decide who gets the job. These algorithms can reflect existing injustices and #discrimination in our #society. We want to fight against this!"*

The posts finished with a call to action and the link to the data donation website: "Help us and our partners from all over Europe to prevent discrimination through algorithms in job allocation. We need real CVs for this! Donate here [Link]."

All consortium partners shared the information and called to donate via email in their professional networks. The project's newsletter was also employed to advertise the donation campaign. In addition, project partners included the data donation in their talks and keynotes at conferences. Along this line, the partners published institutional press releases about the

donation campaign, and funds were invested in advertising the donation campaign on LinkedIn.

To reach historically marginalized communities, which are usually underrepresented in data sets, a collaboration was initiated with a consortium partner that has well-established connections with initiatives and groups focusing on marginalized communities. We engaged in community meetings and disseminated information regarding the data donation campaign, thereby enabling the community members to articulate their concerns and formulate inquiries.

Results

The data donation campaign started in June 2023 and remained open until the end of May 2024. In total, 1143 donations were received, four of which were discarded because of incomplete fields. The remaining 1139 complete submissions included 1211 CVs, considering that about 15% of them contained two CVs, and 7% did not attach documents. Most submissions were in Spanish (78%, 895 out of 1139), followed by English (12%, 132 out of 1139). A similar pattern can be found in the language distribution of CVs, where Spanish leads with 69% of the CVs (836 out of 1211), being English the second most frequent language (19%, 227 out of 1211).

As depicted in Figure 2, half of the donors declared themselves as professionals (567 out 1139) from various sectors, including science and engineering, business and administration, ICT, legal, social, cultural, and health. Other large groups of donors include clerical support workers (14%, 164 out of 1139), such as customer services clerks, general and keyboard clerks, numerical and material recording clerks, and service and sales laborers (12%, 136 out of 1139), including personal, protective, and care services. To a lesser extent, we received donations from managers (121), laborers in elementary occupations (62) like mining, manufacturing, or construction, and craft and related trade workers (51).



Figure 2. Frequency distribution of donors' professional characteristics

The data show a balance between junior and senior workers. One-third of donations belong to juniors with less than five years of professional experience (365 out of 1139), and about the same number of donations are from seniors with 15 years of experience or more (371 out of 1139). Regarding age, 50% of the donors are between 26 and 45 years of age (581 out of 1139), as shown in Figure 3.



Figure 3. Frequency distribution of donors' age

About 50% of the donations were submitted by women (576 out of 1139); from the rest, 44% corresponded to men (502 out of 1139). About 20% of donors reported belonging to the LGBTQ+ community (192 out of 1139), a similar proportion declared being part of a minority group (229 out of 1139), and 15% perceived themselves as foreign in the country where they live (179 out of 1139).

Regarding religion/belief, about 45% of the campaign participants (491 out of 1139) reported being either secular or not religious. Also, Figure 4 shows that almost 40% (420 out of 1139) presented themselves as Christians, while Muslims, Buddhists, Hinduists, and Jews are marginally represented in the data. Less than 10% of donors (84 out of 1139) declared having a disability.



Figure 4. Frequency distribution of donors' religion/belief

In summary, the donation campaign successfully collected a representative and medium-sized dataset composed of 1139 submissions that provide valuable insights into a range of professional and demographic characteristics. Donations, primarily from Spanish speakers and professionals across various fields, show a balanced representation of junior and senior workers, a nearly equal gender distribution, and noticeable inclusivity of LGBTQ+, minority, and foreign-born groups. The religious and belief data, as well as disability information, further enrich the dataset, making it a valuable resource for future research in the field of technology-assisted hiring processes.

Discussion

In this paper, a data donation approach is suggested to develop a representative data set as a basis to address the lack of representativeness in algorithmic hiring. Even though we successfully achieved a representative data set of CVs through a data donation, we encountered challenges on the way. Here, asking the right questions, setting safeguard

protocols in place to protect personal information, and achieving representativeness and diversity in the data are discussed.

Sensitive data: risks and risk mitigation

As described in the method section, we placed great emphasis on inclusive questioning to ensure that as many people as possible felt comfortable answering the sensitive data questions. Apart from the best way to ask for sensitive data, an ongoing debate is whether to ask for sensitive data at all. According to Article 9 of the GDPR the use of sensitive data is—in principle—prohibited. However, Article 9(2) names exceptions to this prohibition, like explicit consent to the data collection or research purposes (Bekkum & Borgesius, 2023) both of those conditions apply to our data donation.

Nevertheless, there are legitimate risks and concerns related to sensitive data collection to control against discriminating effects of algorithm-based systems: i) the collected data could be abused or used for other purposes than indicated; ii) data breaches could happen, and iii) the collection of sensitive data does not automatically debias the algorithmic based system— that is a challenging task itself (Bekkum & Borgesius, 2023).

To mitigate those risks and ensure a safe experience for the data donors, we set the following safeguards: i) we explicitly asked for consent and explained, in an understandable way, what the data will be used for; ii) we complied with the GDPR by enabling the withdrawal of personal data at any time of the project; iii) we used the data only for claimed purposes; iv) the data donation campaign, including the collection of sensitive data, underwent ethical approval before starting the data donation campaign.

An essential goal of the data donation was to reach marginalized and historically discriminated groups to create a representative dataset. We recognize that asking marginalized groups for sensitive data is a particularly vulnerable process with the additional risk that through the association between sensitive data and the donated CV, the donors might be

identified as belonging to a minority group. In this sense, we anticipate that researchers who have direct access to the data can associate the CVs with the additional data. Therefore, the risk is further mitigated by limiting access to the data on a need-to-know basis through the legal protections of the Non-disclosure Agreement.

Furthermore, there is a risk of re-identification through the synthetically generated CVs, which could be used to establish an association between a donated CV and sensible data. However, we value that this risk could generally be considered minimal because data intruders would need to find the original set of CVs to be able to reason about them based on the synthetic data, and neither the identity of the donors nor the corpus of donated CVs will be available online. Second, even if the original set of CVs were leaked, when receiving a synthetic CV, a data intruder will most likely not know which pieces of information come from which original CVs.

Additionally, the EU adopted the AI Act in the course of the data donation campaign, which clarifies the conditions under which sensitive data can be collected (EU AI Act, Art. 10). If the collection of sensitive data is strictly necessary to ensure less biased algorithmic based systems one "may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons." (EU AI Act, Art. 10). The sole purpose of collecting the sensitive data was to use it to create a representative data set to develop algorithm-based systems supporting the reduction of discrimination in application processes. In that sense, collecting sensitive data is urgent if we want to ensure equal opportunities (ENAR, 2016).

Donation representativeness

An important aim of the data donation process was to create a representative data set. However, representativeness is one of the biggest challenges in data donation campaigns itself (Bietz, Patrick, and Bloss, 2019). The objective was to engage a large and demographically diverse group of participants, thereby ensuring the dataset is useful for generating representative synthetic data. In this effort, effective communication and promotional strategies, in conjunction with clear messaging regarding the campaign's scope and purpose, play a crucial role in fostering trust and encouraging participation, ultimately contributing to the attainment of the desired representative data set.

However, how representative are donations? It turned out that although most donors are Spanish nationals, the donation dataset is reasonably representative of the European active population. In terms of gender, females represent 51% of the European population (United Nations, 2024), which is in line with the 50% of donations that came from participants who self-identified as women. Regarding age, 50% of donors are between 26 and 45, while in Europe, people between 26 and 45 years of age represent 43% of the population (United Nations, 2024). About 15% of donors reported living in a country different from where they were born, similar to what can be found in the general population, where 12.4% are foreign (Eurostat, 2023).

The donation dataset, however, shows an overrepresentation of certain groups. For instance, while the LGBTQ+ community constitutes around 7% of the population in countries like Spain, Germany, and the UK (Arora, 2024), 20% of donors identified as LGBTQ+. Similarly, ethnic minorities make up 3% of the European population (European Union, 2019), yet 20% of our participants belong to this group. People with disabilities are also overrepresented, comprising 10% of donors compared to 2% in Europe (European Union, 2019). This strong interest may stem from documented discrimination in algorithmic hiring practices affecting these groups (Yam and Skorburg, 2021; Bertrand and Mullainathan, 2003; Tilmes, 2022), as well as targeted outreach efforts, including partnerships with LGBTQ+ and minority organizations and inclusive communication strategies.

Most Europeans are Christians (64%), while agnostic and atheists account for 27% (European Commission, 2019). The opposite is found in the dataset of donations where agnostic/secular represent 43% of the participants and Christians one-third of them. The group of Muslims is, however, fairly representative, accounting for 2% of the general population (European Commission, 2019) and 3.4% in the donation dataset. Other religions, like Buddhists, Jews, and Hinduists, are represented in both the general population and donations with less than 1% (European Union, 2019).

Highly qualified professional sectors (e.g., ICT, engineering, business administration) overrepresent donations. This might be related to the digital means used to run and advertise the campaign, which might facilitate donations from some sectors, i.e., those using computers to support their daily tasks, but complicates participation from sectors that typically involve manual labor (e.g., manufacturing, construction, maintenance). Another explanation might be that the overrepresented sectors are aligned with professionals familiar with using CVs and job search platforms to access the labor market. In this sense, previous research has highlighted the limited access to online hiring tools among workers from the personal care sector when seeking employment opportunities (Rosales *et al.*, 2024). Our distribution of professional sectors is not representative of the European workforce, which is dominated by people in manufacturing, retail, healthcare, and construction (Eurostat, 2008).

Information extraction for synthetic data generation

Besides the previously discussed challenges of handling sensitive data and ensuring representativeness, the information that we can extract from donated data (educational background, work experience, skills, and so on) proved suitable for generating a synthetic dataset of CVs that realistically mimics the attributes of various groups of workers. The generation of synthetic CVs will be conducted following an approach that combines manual

and automatic tasks. Figure 7 (see Appendix 3) presents a schematic illustration of the process that includes the manual anonymization of donated CVs, removing all personally identifiable information, the extraction of the free text of CVs into structured data, the shuffling of the structuring data to avoid reidentification, and the generation of statistical distributions and hand-made rules to create synthetic documents that resemble the characteristics of the donated materials. For an outline of the synthetic data generation process, please refer to Appendix 3. The implementation of the generative procedure for synthetic data has not been without its technical challenges and limitations, with the main one being difficulties in information extraction stemming from the diversity of CV formats and layouts. This required manual inspection to identify formats and additional manual processing for the less common ones. Another significant challenge was the language of the CVs, as the majority were not written in English. This complicated text processing tasks and necessitated translating non-English CVs into English, a process that risked losing the nuances and particularities of the original text. Additionally, the length of some CVs, which range from 5 to 30 pages, imposed further difficulties, as text processing tools have limits on the amount of text they can handle. Despite these challenges, the first generation attempt successfully produced thousands of synthetic CVs that closely resembled the demographics and professional characteristics of the original donors.

Future Research

Although our study yielded encouraging outcomes in utilizing data donation to construct a substantial and relatively representative corpus of data, it also highlighted several unresolved inquiries that present avenues for future research. Our experience demonstrated the considerable challenges, time demands, and resource requirements associated with the proper organization, preparation, and execution of a data donation campaign. The complexity mentioned above renders the method impractical in instances where frequent data collection is

necessary, particularly when the characteristics of the target population undergo rapid change. Further research is needed to identify ways to streamline and optimize this method for regular, periodic use.

Furthermore, additional research could examine alternative methods for the respectful, inclusive, and appropriate collection of sensitive data for repeated use. In this regard, a crucial question is how to effectively engage communities less familiar with digital technologies, such as professional sectors that do not typically utilize CVs or online platforms. The experience in Brazil and India, including face-to-face procedures in the process of collecting WhatsApp data, may offer a valuable model to explore in future campaigns (Garimella and Chauchard, 2024).

Conclusion

The use of algorithmic or AI-based tools is becoming increasingly prevalent, as evidenced for example by their growing presence in the hiring process. Despite the optimistic forecasts, a considerable number of discriminatory outcomes have been documented. A significant factor contributing to discriminatory outcomes resulting from AI-based tools is the lack of representativeness in the data utilized for tool training. To address the issue of lack of representativeness in algorithmic hiring, we employed a data donation strategy to develop a representative dataset. Nevertheless, this approach presents its own set of challenges. For a data donation to be successful, several factors must be considered. Firstly, it is essential to establish trust between the donor and the recipient research team. Secondly, the donor's privacy and transparency, data usage and access, and finally, the necessity to reach and engage a representative set of donors to obtain their data. These challenges are especially pertinent in the context of our objective to engage with marginalized communities. To address these challenges, we implemented a multifaceted approach comprising clear and straightforward communication supported by visualizations to ensure transparency regarding the research

objectives, societal relevance, data handling practices, and privacy measures. Additionally, an ethics review was conducted.

In sum, the data donation campaign was successful in achieving its objective of collecting a representative dataset, particularly in reaching marginalized groups who are often highly underrepresented in data sets. Nevertheless, further research is necessary to enhance the efficacy of data donation campaigns and optimize their methodologies.

Appendix

Appendix 1: Landing Page Data Donation

This section of the project website functioned as the campaign's landing page. Located in the upper right corner is a link to the frequently asked questions (FAQs) section. Additionally, a link to a webpage outlining the project's privacy procedure can be found in the upper right corner.



Figure 5. Landing page used in the data donation campaign

Appendix 2: Data Donation Survey

The following screenshots depict different sections of the survey used for the collection of

CVs and sensitive information about donors.

DATA DONATIO	DN CAMPAIGN	
Help us mak	ke hiring fair	
	•	
You are invited to donate your CVs through the web for	orm below. Please donate a CV that you would no	ot
feel uncomfortable posting online, and that does not o	contain identifiable information, such as name,	
nome address, phone number, email address, and per personal information on your CV we will aponymize it	rsonal picture. If you accidentally leave some	
anonymization procedure can be found here).	to protocycal princy (teeninear details of the	
Before concluding, you will be asked to optionally pro-	wide some sensitive information including	
berere concluding, you will be daked to optionally pro	was some venarive information, including	
ethnic/cultural background and religion. This data lets	s us examine whether automated hiring systems	
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you	s us examine whether automated hiring systems choose to share this data, it will be protected,	
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o	s us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy	
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o commitment form.	s us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy	
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o commitment form.	us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy CV Language *	
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o commitment form. cv* Choose File	sus examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy cy Language * Select one option	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, adaccessible only by a limited number o commitment form. cv* Choose File No file chosen	su se xamine whether automated hiring systems choose to share this data, it will be protected, fresearchers, who will sign a data privacy cV Language * Select one option Please select the language of the CV	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o commitment form. Cy* Choose File No file chosen Please attach a CY file in PDF, DOC, or ODT. Remember to resource all dering hole more thome advances of the	s us examine whether automated hiring systems choose to share this data, it will be protected, fresearchers, who will sign a data privacy of Language * Select one option Please select the language of the CV	×
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number o commitment form. Cy* Choose File No file chosen Please attach a CV file in PDF, DOC, or OOT. Remember to remove all identifiable information (name, home address, phone number, email address, and/or persona joitumi before	us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy cV Language * Select one option Please select the language of the CV	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number of commitment form. CV* CV* Choose File No file chosen Please attract AV file in PBF, DOC, or DOT. Remember to remove al identifiable information (name, home address, phone number, email address, and/or person joicture) before uploading. File must be of 50MB of size at maximum.	us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy CV Language * Select one option Please select the language of the CV	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number of commitment form. CV* Choose File No file chosen Please attach a CV file in PDF, DOC, or OOT. Remember to remove al luderable information (name, home address, phone number, amil address, and/or personal picture) buffer updadage. File multi-be of 500H of size at maximum. CV in an alternate lenguage	sue seamlne whether automated hiring systems choose to share this data, it will be protected, fresearchers, who will sign a data privacy CV Language * Select one option Please select the language of the CV Alternate CV Language	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number of commitment form. CV* Choose File No file chosen Please attach a CV file in PDF, DOC, or OOT. Remember to remove al ledinative information (name, home address, phone number, email address, andre presonal picture) before updading. File multi be of 50MI of a set maximum. CV is an alternate lenguage Choose File	sue seamlne whether automated hiring systems choose to share this data, it will be protected, fresearchers, who will sign a data privacy CV Language * Select one option Please select the language of the CV Atternate CV Language Select one option	~
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, adaccessible only by a limited number of commitment form. CV* CV* Choose File No file chosen Please attach a CV file in PDF, DOC, or ODT. Remember to remove all identifiable information (name, home address, phone number, email address, andro personal picture) before upolading. File must be of SOM Bor state at maximum. CV is an alternate language Choose File No file chosen	s us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy cV Language * Select one option Please select the language of the CV Alternate CV Language Select one option Please select the language of the alternative CV	v
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number of commitment form. CV* C* C* C* C* C* C* C* C* C* C	us examine whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy of Language * Select one option Please select the language of the CV Alternate CV Language Select one option Please select the language of the alternative CV	v
ethnic/cultural background and religion. This data lets might discriminate against certain job seekers. If you encrypted, and accessible only by a limited number of commitment form. CV* CV* Choose File No file chosen Please attract AC VIIe in PDF, DOC, or ODT, Remember to remove al identifiable information (name, home address, phone number, email address, androp remove joicumb Hofro uploading. File must be of 50MB of size at maximum. CV or an alternate language Choose File No file chosen Optionally, you can donate a CV equivalent to the previously attached, but in a different language. Remember to remove all interfitiable information before uploading. File must be of 50MB	sue seamlne whether automated hiring systems choose to share this data, it will be protected, f researchers, who will sign a data privacy of Language * Select one option Please select the language of the CV Alternate CV Language Select one option Please select the language of the alternative CV	~

Lidon't have a CV	
or can also participate If you do not have a CV. Please fill in manually: your educational background, professional experience, a skills, such as the languages you speak.	and
Job category (first level) *	
Select one option	~
Please indicate the type of job that the donated CV is relevant for	
Job category (second level) *	
Select one option	~
Please indicate the type of job that the donated CV is relevant for	
Years of professional experience *	
Select one option	~
Please choose the years of professional experience (including internships) that the donated CV reflects for the indicated job category	
Highest level of education attained *	
Select one option	~
Please choose the highest level of formal education you have attained	
How old are you? *	
Prefer not to answer	~
Please select your age range	

a) The section to submit CVs

What gender do you identify with?		
Prefer not to answer		~
Please select whichever option is most comfortable for you.		
Does your gender differ from the gender assigned at birth?		
Prefer not to answer		~
Please indicate whether you identify with a gender different fror	n the one assigned at birth. Select one option	
What is your sexual orientation?		
Prefer not to answer		~
Please select whichever option is most comfortable for you.		
	Libre 7	
Are you perceived as "toreign" or not "white" in the country where you	live?	
Are you perceived as "foreign" or not "write" in the country where you Prefer not to answer Racial discrimination is often experienced by people who are pe	Inve? rceived as "foreign" or not "white".	~
Are you proteived as "toreign" or not "write" in the country where you Prefer not to answer Racial discrimination is often experienced by people who are pe independently of what you answered to the previous question, would you self-describe as belonging to an ethnic minority in the construct whom were the	inve? rceived as "foreign" or not "white". Ethnic group(s)	~
Are you proteived as increagen or not writter in the country where you Prefer not to answer Racial discrimination is often experienced by people who are pe independently of what you answered to the previous question, would you self-describe as belonging to an ethnic minority in the country where you live? Prefer not to answer	reelved as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you	~
Are you proteived as increage or not where in the country where you Prefer not to answer Racial discrimination is often experienced by people who are pe independently of what you answered to the previous question, would you aelf-describe as belonging to an ethic minority in the country where you live? Prefer not to answer Prefer not to answer Prefer not to answer Prefer not to answer Such as cultural identity, origin, first language or accent, affirement rom the majority group. Please select the option that best describe your situation.	inve? rceived as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you belong to. Use commas to separate the groups.	~
Are you proteived as increage or not where in the country where you Prefer not to answer Racial discrimination is often experienced by people who are per independently of what you answered to the previous question, would you self-describe as belonging to an ethic minority in the country where you live? Prefer not to answer Prefer not to answer Prefer indicators as group of people sharing factors such as cultural identity, origin, first language or accent, afferent from the majority group. Please select the option that best describe your situation. What religious or philosophical community do you belong to?	Inver rceived as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you belong to. Use commas to separate the groups.	~
Ne you proteive as in tereign on the where in the country where you Prefer not to answer Racial discrimination is often experienced by people who are people independently of what you answered to the previous question, would you self-decembe as belonging to an ethnic minority in the country where you live? Prefer not to answer V Think minority decembes a group of people sharing factors such as cultural identity, origin, first language or accent, different from the majority group. Please select the option that best describe your situation. What neligious or philosophical community do you belong to? Prefer not to answer	Inve? rceived as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you belong to. Use commas to separate the groups.	~
Ne soa protevende as torelegin or hot where in the country where you Prefer not to answer Racial discrimination is often experienced by people who are per independently of what you answered to the previous question, would you self-describe as belonging to an ethic minority in the country where you live? Prefer not to answer Function minority describes a group of people sharing factors such as cultural identity, origin, first language or accent, afferent from the majority group. Please select the option that best describe your situation. What neligious or philosophical community do you belong to? Prefer not to answer Please select your religion or belief	Inver rceived as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you belong to. Use commas to separate the groups.	~
Ne so a processe as increage or not where in the country where you Prefer not to answer Racial discrimination is often experienced by people who are per independently of what you answered to the previous question, would you self-describe as belonging to an ethic minority in the country where you live? Prefer not to answer Function minority describes a group of people sharing factors such as cultural identity, origin, first language or accent, afferent from the majority group. Please select the option that best describe your situation. What neligious or philosophical community do you belong to? Prefer not to answer Please select your religion or belief Do you have a disability?	Inver rceived as "foreign" or not "white". Ethnic group(s) You may optionally list one or more of the ethnic groups you belong to. Use commas to separate the groups.	~

b) To indicate professionally-related attributes

c) This section contained sensitive data questions

Figure 6. Online survey used to collect donations

Appendix 3: Synthetic CVs Generation

Synthetic data is data artificially generated through a deliberate procedure to address data-related tasks (Jordon *et al.*, 2022; Ortega, Bourgeois and Kortuem, 2021). It has become a widely used method for augmenting existing datasets or creating new ones when real records are impossible due to logistical or ethical constraints (Whitney and Norman, 2024). This is the case of research on fairness in algorithmic hiring, which, due to its nature, needs to deal with sensitive demographic data about job seekers. In this context, synthetic data has been broadly recognized for its value as a method to avoid privacy concerns and reduce bias in AI-based algorithms by removing imbalances and suppressing disparate impact (Fabris *et al.*, 2023).

In our case, the generation of synthetic CVs is proposed to be conducted following an approach that combines manual and automatic tasks. Figure 7 presents a schematic illustration of the process that starts with manually anonymizing donated CVs removing all personally identifiable information (step 1 in Figure 7). Even though donors are asked to delete personal information in CVs, this step ensures that no personal data is contained in donated materials. It was decided to be conducted manually because none of the automatic tools for removing personal information from free text achieved decent results for all CV formats. Later, anonymized CVs are converted into structured documents (step 2 in Figure 7). To avoid associations between the collected sensitive data and CVs, which may lead to identifying a person as part of a minority group, the content of structured CVs is shuffled (step 3 in Figure 7). Even when CVs are previously anonymized, we strived to guarantee that synthetic data cannot be used for re-identification. To ensure this, the main idea is to introduce as much variability as possible in the synthetic CV generator by shuffling donor education institutions and workplaces with similar demographic characteristics. Finally, shuffled CVs are used to generate the synthetic CVs using statistical distributions and hand-made rules to resemble the

characteristics of the reference CVs and improve the coherence of the created documents (step 4 in Figure 7).



Figure 7. A schematic high-level overview of the approach proposed to generate synthetic CVs.

As a result of the entire process, we ended up with a dataset of synthetic CVs that realistically mimic the attributes of different groups of workers. However, without the risk that the data donors can be identified. In this way, realistic data sets can be created to preserve privacy.

Reference

Araujo, T. et al. (2017) 'How Much Time Do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use', Communication Methods and Measures [Preprint]. Available at: https://doi.org/10.1080/19312458.2017.1317337.

Arora, P. (2024) 'Rainbow Operations: Let's Add LGBTQ+ Colors to "Doing Good with Good Operations", Production and Operations Management [Preprint]. Available at: https://doi.org/10.1177/10591478241239007.

Badgett, M.V. (2009) 'Best practices for asking questions about sexual orientation on surveys'.

Baumann, A.-L., Egenberger, V. and Supik, L. (2018) 'Erhebung von Antidiskriminierungsdaten in repräsentativen Wiederholungsbefragungen. Bestandsaufnahme und Entwicklungsmöglichkeiten', Antidiskriminierungsstelle des Bundes: Berlin, Germany [Preprint].

Bertrand, M. and Mullainathan, S. (2003) xr. w9873. National Bureau of Economic Research. Available at: https://doi.org/10.3386/w9873.

Rosales *et al.* (2024) Beyond the Algorithm: Expanding the understanding of fairness and non-discrimination in algorithmic hiring – The case for Latin American migrants seeking employment opportunities in Spain AI Advisory Boards. Available at:

https://aiadvisoryboards.wordpress.com/2024/01/27/beyond-the-algorithm-expanding-the-understandi ng-of-fairness-and-non-discrimination-in-algorithmic-hiring-the-case-for-latin-american-migrants-see king-employment-opportunities-in-spain/ (Accessed: 17 December 2024).

Boeschoten, L. et al. (2022) 'Privacy-preserving local analysis of digital trace data: A proof-of-concept', Patterns (New York, N.Y.), 3(3). Available at: https://doi.org/10.1016/j.patter.2022.100444.

Chander, S. (2017) 'Women of Color in the Workplace'.

Christner, C. et al. (2022) 'Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations', Communication Methods and Measures [Preprint]. Available at: https://doi.org/10.1080/19312458.2021.1907841.

corporate-body. JUST:Directorate-General for Justice and Consumers, N.A.K. (2019) Discrimination in the European Union. Publications Office of the European Union.

Couto, Garimella (2024) 'Examining (Political) Content Consumption on Facebook Through Data Donation'. Available at:

https://cs.paperswithcode.com/paper/examining-political-content-consumption-on (Accessed: 17 December 2024).

van Driel, I.I. et al. (2022) 'Promises and Pitfalls of Social Media Data Donations', Communication Methods and Measures [Preprint]. Available at: https://doi.org/10.1080/19312458.2022.2109608.

EIGE (2021) Covid-19 derails gender equality gains. Available at: https://eige.europa.eu/newsroom/news/covid-19-derails-gender-equality-gains.

ENAR (2016) Towards LGBTIQ Equality: Free To Be Yourself In The EU. Available at: https://www.enar-eu.org/wp-content/uploads/edc-general factsheet final.pdf.

European Commission (2019) Final report of the High-Level Expert Group on the Impact of the Digital Transformation on EU Labour Markets. Available at:

https://digital-strategy.ec.europa.eu/en/news/final-report-high-level-expert-group-impact-digital-transf ormation-eu-labour-markets (Accessed: 3 January 2025).

Europen Commission (2020) Towards Lgbtiq Equality: Free To Be Yourself In The EU. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0698.

Eurostat (2023) Migration and asylum in Europe – 2023 edition.

Fabris, A., Dennis, M.J. and Graus, D. (2024) Fairness and bias in algorithmic hiring: A multidisciplinary survey, aithor.com. Available at: https://aithor.com/paper-summary/fairness-and-bias-in-algorithmic-hiring-a-multidisciplinary-survey (Accessed: 17 December 2024).

Michèle Finck, Frank Pallas, They who must not be identified—distinguishing personal from non-personal data under the GDPR, *International Data Privacy Law*, Volume 10, Issue 1, February 2020, Pages 11–36, https://doi.org/10.1093/idpl/ipz026 (Accessed: 03 January 2025).

Garimella, K. and Chauchard, S. (2024) 'WhatsApp Explorer: A Data Donation Tool To Facilitate Research on WhatsApp'. Available at: http://arxiv.org/abs/2404.01328 (Accessed: 17 December 2024).

Hummel, P., Braun, M. and Dabrock, P. (2019) 'Data Donations as Exercises of Sovereignty', The Ethics of Medical Data Donation, pp. 23–54.

Keusch, F. et al. (2024) 'Do You Have Two Minutes to Talk about Your Data? Willingness to Participate and Nonparticipation Bias in Facebook Data Donation', Field Methods [Preprint]. Available at: https://doi.org/10.1177/1525822X231225907.

Kovacheva, S. et al. (2018) 'Growth, Equal Opportunities, Migration and Markets (GEMM) Report on institutional and contextual factors'.

Kuiper, K. (2021) Developing a user interface design to foster trust and maximize consent for data donation in scientific research. Available at: https://studenttheses.uu.nl/handle/20.500.12932/293 (Accessed: 18 December 2024).

Loecherbach, F. (2022) OSD2F: An Open-Source Data Donation Framework, Personal Website Felicia Loecherbach. Available at: https://www.felicialoecherbach.com/publication/osd2f/ (Accessed: 18 December 2024).

Ohme, J. et al. (2020) 'Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function', Mobile Media & Communication [Preprint]. Available at: https://doi.org/10.1177/2050157920959106.

Ohme, J. et al. (2024) 'Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking', Communication Methods and Measures [Preprint]. Available at: https://doi.org/10.1080/19312458.2023.2181319.

Ortega, A.G., Bourgeois, J. and Kortuem, G. (2021) 'Towards Designerly Data Donation', in UbiComp/ISWC 2021 - Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers. 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2021 ACM International Symposium on Wearable Computers, UbiComp/ISWC 2021, Association for Computing Machinery (ACM), pp. 496–501.

Patterns (2022) 'Digital data donations: A quest for best practices', Science Direct. Available at: https://doi.org/10.1016/j.patter.2022.100467

Palmer, M. and Harley, D. (2012) 'Models and measurement in disability: an international review', Health policy and planning, 27(5), pp. 357–364.

Portela, M. et al. (2024) 'A comparative user study of human predictions in algorithm-supported recidivism risk assessment', Artificial Intelligence and Law, pp. 1–47.

Puschmann, C. (2019) 'Beyond the Bubble: Assessing the Diversity of Political Search Results', Digital Journalism [Preprint]. Available at: https://doi.org/10.1080/21670811.2018.1539626.

Weitzman et al. (2010)'Sharing Medical Data for Health Research: The Early Personal Health Record Experience' Journal of Medical Internet Research, 12(2). Available at: https://doi.org/10.2196/jmir.1356.

Skatova, A. and Goulding, J. (2019) 'Psychology of personal data donation', PLOS ONE, 14(11), p. e0224240.

Strotbaum, V. et al. (2019) 'Your data is gold – Data donation for better healthcare?', it - Information Technology, 61(5-6), pp. 219–229.

Tilmes, N. (2022) 'Disability, fairness, and algorithmic bias in AI recruitment', Ethics and Information Technology, 24(2), pp. 1–13.

Whitney, C.D. and Norman, J. (2024) 'Real risks of fake data: Synthetic data, diversity-washing and consent circumvention', in The 2024 ACM Conference on Fairness, Accountability, and Transparency. FAccT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency, New York, NY, USA: ACM. Available at: https://doi.org/10.1145/3630106.3659002.

Yam, J. and Skorburg, J.A. (2021) 'From human resources to human rights: Impact assessments for hiring algorithms', Ethics and Information Technology, 23(4), pp. 611–623.