Location matching on shaky grounds: Re-evaluating algorithms for refugee allocation

Clara Strasser Ceballos
LMU
Munich, Germany
Clara.StrasserCeballos@stat.uni-muenchen.de

Christoph Kern
LMU
Munich, Germany
Munich Center for Machine Learning (MCML)
Munich, Germany
christoph.kern@stat.uni-muenchen.de

Abstract

The initial location to which refugees are assigned upon arrival in a host country plays a key role in their integration. Several research groups have developed tools to optimize refugee-location matching, with the overall aim of improving refugees' integration outcomes. Four primary tools are already being piloted across various countries: GeoMatch, Annie™ Moore, Match'In, and Re:Match. The first two tools combine supervised machine learning with optimal matching techniques, while the latter two rely on heuristic methods to match refugee preferences with suitable locations. These tools are used in a highly sensitive context and directly impact human lives. It is, therefore, not only desirable but critical to (re-)evaluate them through the lens of algorithmic fairness. We contribute in three key aspects: First, we provide a comprehensive overview and systematization of the tools aimed at the algorithmic fairness community. Second, we identify sources of biases along the tool design stages that can contribute to disparate impacts downstream. Finally, we simulate the application of the GeoMatch tool using German survey data to empirically illustrate the impact of target variable choice on matching outcomes. While GeoMatch optimizes economic integration, we demonstrate that the integration gains differ substantially when social integration is prioritized instead. With our use case, we highlight the susceptibility of algorithmic matching tools to design decisions such as the operationalization of the integration outcome and emphasize the need for more holistic evaluations of their social impacts.

CCS Concepts

• **Computing methodologies → Machine learning**; • **Applied computing → Law, social and behavioral sciences**.

Keywords

matching tools, integration, refugees, fairness evaluation

ACM Reference Format:
Clara Strasser Ceballos and Christoph Kern. 2025. Location matching on shaky grounds: Re-evaluating algorithms for refugee allocation. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3715275.3732149

FAccT '25, Athens, Greece

ACM ISBN 979-8-4007-1482-5/25/06
https://doi.org/10.1145/3715275.3732149

1 Introduction

Wars, violence, political persecution, famine, and other destructive events force people to flee their homes in search of refuge in new host societies. According to the United Nations High Commissioner for Refugees (UNHCR), around 2 million refugees needed resettlement in 2023 [119]. However, less than 5% of those in need successfully resettle to third countries [46, 119]. Upon arrival in the host countries, the locations to which refugees are eventually assigned depend on the legal and administrative framework in place, with decisions being largely influenced by location-specific constraints [99]. In some countries like Switzerland and Germany, location allocations are even made (quasi-)randomly [12, 40]. As a result, most allocation processes fail to incorporate refugees' characteristics and preferences [5]. Yet, research has shown that initial placement plays a critical role in the integration outcomes for refugees [5, 8]. To address this gap, several research groups have developed tools that take refugees' characteristics and preferences into account when allocating refugees, with the aim of improving their integration outcomes. These tools are: GeoMatch, Annie™ Moore, Match'In and Re:Match[1] [5, 13, 103, 112].

The first two tools, GeoMatch and Annie™ Moore, assign refugees to locations based on predictions of integration outcomes [5, 13]. In short, the tools train machine learning models on historical data containing information on refugee characteristics, assigned locations, and specific measures of integration (e.g., employment status). Integration predictions are generated for newly arriving refugees. These refugees are assigned to locations that maximize an optimality criterion (e.g., global average employment) subject to constraints (e.g., location capacity). These tools are currently being piloted by two U.S. resettlement agencies, Global Refuge, and HIAS, as well as in Switzerland and the Netherlands [59]. Match'In and Re:Match, in contrast, assign refugees to locations that best match their preferences and needs [103, 112]. Hence, these tools do not rely on predictive modeling but make use of real-time survey data on refugees' preferences and corresponding information about locations. Matches are generated based on pre-defined rules and similarity measures and then ranked accordingly. Both tools are being piloted in Germany [86, 101].

As these tools directly impact vulnerable populations, it is critical to carefully assess potential sources of disparate impact and threats to reliability. Yet, location matching tools present a unique type of algorithmic decision-making (ADM) system: 1) there are

[1] The four tools are the only major tools developed and piloted at scale to date. This claim is substantiated by the overviews provided by Ozkul [95], Match'In and Re:Match [86, 101], as well as by our own literature and documentation research.

no inherently correct "ground truth" location assignments against which the algorithms' recommendations can be evaluated, and 2) prediction-based algorithms include both a prediction and a matching layer. This challenges the direct application of key (bias) concepts of algorithmic fairness and, at the same time, elevates the role of decisions taken in the design of the allocation systems. The exact operationalization of the integration outcome to optimize, for example, critically drives the allocation decisions while leaving limited options for assessing their true effectiveness. While research has raised concerns about the ethical aspects of location matching tools [23, 36, 72, 89, 95, 104] and proposed fairness measures for assessing the location assignments [17, 56], to the best of our knowledge, no comprehensive study has been conducted that 1) provides a transparent comparison between the four major matching tools, and 2) systematically studies potential sources of errors and biases along the tools' design stages that can harm their reliability.

This study contributes in three key aspects. First, we provide a detailed and up-to-date overview of the four existing refugee-location matching tools. Second, we highlight key components that must be considered along the design stages of such tools, identifying both common and unique sources of unfairness specifically for location matching algorithms. These components include the data source and collection method, target variable definition and operationalization, feature selection, and the potential for concept drifts. We systematically examine the critical aspects of each component and their interaction with different sources of errors and bias (including historical, representation, and measurement bias, selective labels, selection, and confounding bias) that can lead to reliability issues and reduced tool performance. Finally, we empirically illustrate the importance of tool design decisions – in particular, the operationalization of the target variable – for algorithmic matching outcomes. To this end, we use German refugee survey data to simulate the application of the GeoMatch tool in the German context. We emphasize the importance of the operationalization of the target variable in determining GeoMatch's integration outcome gains.

Based on our analytical findings from the GeoMatch case study, we highlight the importance of design decisions, such as different target variable definitions, on allocations. GeoMatch currently uses employment outcomes as the only measure of refugee integration [13]. However, employment represents only the economic dimension of integration [67]. It is crucial to account for other variables that capture further dimensions of integration. To illustrate this, we construct two target variables: one measuring economic (measured by employment status) and another measuring social integration (measured by the frequency of social activities, including social contacts with natives and non-natives). When the tool assigns refugees to locations to maximize employment outcomes, employment gains reach approximately 98% over random allocation, meaning the employment rate more than doubles with presumed algorithmic compared to actual assignments. However, this approach only achieves a 14% increase in social integration. Conversely, if the tool optimizes for social integration, social integration rates increase by 56%. Based on these results, we recommend agencies that aim to implement algorithmic matching tools to carefully assess, conceptualize, and implement their matching objectives, e.g., by considering indices that incorporate multiple integration dimensions [67].

Our paper is structured as follows. We first provide an overview of the four existing matching tools (Section 2). Further, we outline errors and biases that may arise during the tools' design stage (Section 3). We then present our use case and its results (Section 4) and conclude with a discussion (Section 5).

2 Matching Tools: An Overview

We provide an overview of the four existing refugee-location matching tools[2]. These tools share the same goal: to improve the integration of refugees by matching them to a presumably optimal location upon arrival in a host country. The tools are designed for refugees without pre-existing family ties in the host country (so-called "free-cases") or those being resettled simultaneously, as family members are typically assigned to the same location. Despite sharing the same aim, the tools differ substantially in design and methodology. To encourage a dialog about the reliability of the tools and to evaluate potential strengths and limitations, a joined overview of the tools is crucial. However, existing research has primarily focused on high-level ethical considerations or broad comparisons [89, 95], leaving a gap in detailed analysis. To address this, we conducted comprehensive research on each tool. This resulted in a transparent comparison of the tools across several key dimensions as summarized in Table 1 and outlined below. Further details on the resettlement process in each country in which the tools are currently being piloted, as well as on the methodology of each tool, can be found in Appendix B and C.

Methodology. To create the overview, we systematically reviewed original research papers, documentation, and supplementary materials related to each tool. Additionally, we analyzed official websites and articles written about the tools. We also searched for available code bases and conducted interviews with project leads.

2.1 GeoMatch

GeoMatch[3], developed by the Immigration Policy Lab (Stanford and ETH Zurich), was first introduced by Bansak et al. [13]. The authors apply the tool in two main contexts: the U.S. and Switzerland. In the U.S., the tool is used by the U.S. resettlement agency Global Refuge[4] to assign refugees to one of its local affiliates upon arrival. The tool aims to optimize refugees' employment outcomes 90 days after arrival by considering their individual-level characteristics in the matching process. Administrative data from Global Refuge is used for this purpose. In Switzerland, the tool is used to match asylum seekers (see Appendix A for definitions of terms used in this paper) to one of Switzerland's 26 cantons. In this context, asylum seekers' employment outcomes two, three, and four years after arrival are sought to be maximized. Data from the ZEMIS database is used [105]. Methodologically, GeoMatch combines supervised machine learning with optimal matching algorithms. The process involves, in brief, the following four steps (see Appendix C). First, historical data is divided into train and test data. Second, a prediction model is fitted on the train data, and employment probability predictions are generated for the refugees in the test data. Predictions are made for each refugee and for each potential location to which they could be assigned. The refugees in the test data are matched such that an

[2]We will often broadly refer to "refugees" throughout this paper. However, it is important to note that the population targeted by these tools may also include asylum seekers (e.g., for Match'In). The distinction between the terms is detailed in Appendix A.
[3]Website: https://immigrationlab.org/geomatch/.
[4]Former name: Lutheran Immigration and Refugee Service (LIRS).

Tool	Pilot Study	Target Population	Location Unit	Data Source (Type)	Variables	Methodology	Tool Extensions and Evaluations
GeoMatch [13–15]	U.S.	**Refugees:** - Age: 18-64	Local affiliates of Global Refuge	Global Refuge (Administrative)	**Features:** - Individual-level characteristics (sex, age at arrival, country of origin, year & month of arrival, free-case, education level, english speaking) **Target:** - Employment 90 days after arrival (no/yes)	**Stages:** **1.)** Data Setup: Train-test split **2.)** Modeling: Fit model and generate predictions (Method: gradient-boosted trees) **3.)** Mapping: Transform individual-level to case-level predictions (Metric: probability that at least one individual finds employment in the location) **4.)** Matching: Match case to location to maximize global average employment, s.t. capacity constraints	- Decision support tool for economic migrants in Canada: [52] - Incorporate (simulated) refugee preferences into the matching stage: [2, 3] - Evaluate fairness of matching: [17, 56] - Consider distributional shifts: [16, 18, 19] - Adapt matching tool to the Netherlands context: [94] - **Evaluation results pilot: NA**
	Switzerland	**Asylum Seekers:** - Age: 18-65 - Residence status: F permit (temporary admission as a refugee)	26 cantons of Switzerland	ZEMIS (Administrative)	**Features:** - Individual-level characteristics (sex, age at arrival, country of origin, year & month of arrival, free-case, marital status, christian, Muslim, french-speaking) **Target:** - Employment 2, 3, 4 years after arrival (no/yes)		
Annie™ Moore [5, 6]	U.S.	**Refugees:** - Age: Working age	Local affiliates of HIAS	HIAS (Administrative)	**Features:** - Individual-level characteristics (gender, age upon arrival, arrival date, relationship status, children, nationality, language, education level, medical condition, treatment urgency, urgency code) - Country-level characteristics (average employment level & average unemployment rate) **Target:** - Employment 90 days after arrival (no/yes)	**Stages:** **1.)** Data Setup: Train-test split **2.)** Modeling: Fit model and generate predictions (Method: LASSO) **3.)** Mapping: Transform individual-level to case-level predictions (Metric: sum of probabilities) **4.)** Matching: Match cases to locations to maximize total expected number of employed refugees, s.t. binary service and capacity constraints	- Incorporate (simulated) preferences into the matching stage: [44, 45] - Consider distributional shifts: [7] - **Evaluation results pilot: NA**
Match'In [103]	Germany	**Asylum Seekers:** - Age: ≥ 18 - No spouse or family - Regular asylum process	Municipalities in Germany	Survey Data	**Features:** - Individual-level characteristics & special protection needs - Individual-level preferences (Areas: social ties, living, work, education, health, leisure time, advice and help) - Location-level characteristics - Location-level capacities (Areas: as individual-level preferences)	**Stages:** **1.)** Evaluation: Assess feasibility of matching **2.)** Case Generation: Generate ideal municipality **3.)** Similarity: Calculate similarity (Method: Case-Based Reasoning) **4.)** Matching: Rank and match based on similarity	- Information letter: [100] - Survey Questionnaire: [24] - **Evaluation results pilot: NA**
Re:Match [112]	Germany	**Refugees:** - Ukrainians fleeing from the Russian war - Valid foreign passport/ID - No temporary residence permit in Germany yet	Municipalities in Germany	Survey Data	**Features:** - Individual-level characteristics & special protection needs - Individual-level preferences (Areas: family/medical support services, cultural support, housing accessibility, professional services) - Location-level characteristics - Location-level capacities (Areas: as individual-level preferences)	**Stages:** **1.)** Evaluation: Assess feasibility of matching **2.)** Similarity: Calculate similarity **3.)** Matching: Rank and match based on similarity	- Recommendation guide: [98] - Implementation guide: [48] - **Evaluation results pilot:** [33, 112]

Table 1: Location matching tools are trialed in multiple contexts, but evaluation results of pilot projects are scarce or largely absent. The table provides an overview of the four existing matching tools across several (design and application) dimensions.

optimality criterion (e.g. global average employment) is maximized while accounting for location-specific constraints. GeoMatch has been piloted by the Swiss State Secretariat for Migration (SEM) since 2020 and by Global Refuge since 2023 [59]. Additionally, the tool has been tested by the Dutch Central Agency for the Reception of Asylum Seekers (COA) in the Netherlands since 2024 [94]. The GeoMatch functionality as presented by Bansak et al. [13] has since then been extended [2, 17–19, 52, 56]. Evaluation results of the pilot projects are not publicly available.

2.2 Annie™ Moore

Annie™ Moore[5], jointly developed by the Universities of Oxford, Lund, and the Worcester Polytechnic Institute, was first introduced by Ahani et al. [5]. Similar to GeoMatch, the tool has been applied in the U.S. context. Specifically, the tool is used by one U.S. resettlement agency, the Hebrew Immigrant Aid Society (HIAS). The tool aims to optimize refugees' employment outcomes 90 days after arrival by considering their individual-level characteristics. Methodologically, Annie™ Moore shares similarities with GeoMatch, combining supervised machine learning with optimal matching techniques. However, it differs in the primary prediction method used and the optimality criteria and constraints chosen for the matching stage (see Appendix C). The tool has been piloted by HIAS since 2018 [93]. Ahani et al. [7] extended the tool to Annie™ Moore 2.0. Evaluation results of the pilot project are not available.

[5]Website: https://www.refugees.ai/.

2.3 Match'In

Match'In[6], jointly developed by the University of Hildesheim and the Friedrich-Alexander University of Erlangen-Nuremberg, was introduced by Sauer et al. [103]. The tool, used in the German context, aims to match asylum-seekers residing in initial reception centers with German municipalities that best match their preferences. For this purpose, survey data is collected from voluntarily participating asylum seekers and municipalities. Unlike the first two tools, Match'In does not rely on supervised learning. Instead, the matching process involves the following steps: evaluating whether a match is feasible based on exclusion criteria, generating an "ideal municipality" for each asylum seeker based on their preferences, using a case-based reasoning (CBR) approach to identify municipalities most similar to the ideal one, and ranking them accordingly. The tool was piloted from 2023 to 2024 in four federal states[7] in Germany [86]. The evaluation results of the pilot project are not yet available[8].

2.4 Re:Match

Re:Match[9], jointly developed by Pairity, Berlin Governance Platform, Right to Protection (R2P) and Salam Lab, was first presented by Smith et al. [112]. The tool, applied in the German context, aims

[6]Website: https://matchin-projekt.de/en/.
[7]Hesse, Lower Saxony, North Rhine-Westphalia, and Rhineland-Palatinate.
[8]The results of the evaluation are intended to be made publicly available according to the tool's website (https://matchin-projekt.de/en/project-results/).
[9]Website: https://rematch-eu.org/.

to match Ukrainian refugees residing in Poland or planning to flee from Ukraine with German municipalities that best match their preferences. To this end, survey data is collected from voluntarily participating refugees and municipalities. Similar to Match'In, the tool does not use supervised learning. The matching process involves the following steps: a feasible set of municipalities is identified for each refugee, and a score is calculated for each municipality based on the match between the refugee's background characteristics and preferences and the municipality's characteristics and capacities and ranked accordingly. The tool was piloted in eight German municipalities across several states in two phases between 2022 and 2024 [48]. Evaluation results (including satisfaction and integration after matching) for the first phase are publicly available [33, 112].

3 Tool Design Components: Mapping Error Sources to Design Decisions

The life cycle of matching tools broadly includes the design, implementation, and evaluation stages. Each stage includes critical components that must be addressed to avoid harm to any (sub)group. In this study, we focus on the design stage, which involves the initial steps taken by researchers and developers to conceptualize and design the matching system, and lays the groundwork for subsequent stages. Unaddressed issues at this stage can critically exacerbate the tool's (disparate) impacts. Accordingly, we examine the following components [53, 110]: the role of the data source and collection method, the definition of the target variable, the selection of features, and the handling/ anticipation of concept drift. We thereby map sources of bias discussed in the algorithmic fairness literature [58, 87, 88, 115] to the matching context, highlighting critical decision points along the tools' design stages.

Since the four matching tools differ in methodology, we categorize them into prediction-based tools (GeoMatch and Annie™ Moore) and non-prediction-based tools (Match'In and Re:Match). The distinction is relevant as design decisions within different methodological frameworks can lead to varying consequences.

3.1 Data Source & Collection Method

Overview. Reliable data is fundamental not only in ensuring accurate models but also in addressing fairness concerns [20]. A detailed overview of biases in data is provided by Mehrabi et al. [87] and Suresh and Guttag [115]. Among these, two key biases are particularly relevant in our context: biases inherent in the data itself and biases arising from the data collection method. First, historical bias is a well-known concern that arises when the data reflects pre-existing biases in the world, such as discrimination based on gender or race [115]. Second, representation bias occurs when certain groups are under (or over) represented in the data, e.g., as a result of an inadequate sampling method [115]. This can result in some groups being more likely to be included in the final data sample than others (selection bias) or in some groups being more likely to self-select into the sample (self-selection bias) in comparison to the composition of groups in the target population [97, 106]. This type of bias is intensively studied in the survey data context [26, 61]. Survey data may additionally suffer from non-sampling errors like measurement error [26] (see also Section 3.2).

Refugee-Location Matching Context. Prediction-based matching tools are subject to historical bias when the optimization outcome in the training data is affected by discrimination processes and social inequalities. In the context of refugee resettlement, the outcome typically represents a (labour market) integration measure, such as whether the refugee finds employment within a given time period [13]. However, these outcomes can reflect underlying discrimination in hiring practices, both between and within different (refugee) subpopulations. Fossati et al. [55], for instance, find that female refugees without children are preferred over male refugees and female refugees with children in hiring decisions across Germany, Austria, and Sweden. Van der Zwan and Van Tubergen [125] demonstrate that Muslim refugee women wearing veils are disadvantaged compared to non-Muslim refugee women in the Netherlands.

Representation bias can arise when the sampling method is nonrandom. In the context of refugee resettlement, two primary data sources are used: administrative and survey data. Here, administrative data refers to data collected by official resettlement agencies or federal offices of all resettled refugees. Using complete data sets or random samples of these records for the matching tools reduces the risk of representation bias. However, administrative data may still suffer from such bias if resettlement offices determine which refugee groups are first distributed and included in their data base. The second type of data source, survey data, refers to information collected for (non-prediction-based) matching projects through surveys. Representation bias may arise if only certain locations or refugee groups are included during the project's recruitment phase (selection bias) or if survey non-participation is systematic (self-selection bias). As a result, refugees and locations that participate may share certain characteristics (e.g., higher language skills, stronger motivation to integrate or better resources, more favorable attitudes towards refugees).

In addition to representation bias, the survey data may suffer from measurement error for a number of reasons [74]: 1) the questionnaire design, e.g., if the survey questions were designed with specific refugee groups in mind [114], 2) the mode of data collection, e.g., if the presence of interviewers influences refugees' responses due to social desirability [69], 3) interviewer characteristics, e.g., if female interviewees adapt their answers when interviewed by men [83], and 4) respondent characteristics, e.g., if respondents misunderstand questions due to their cultural background [114]. Furthermore, as highlighted by Ahani et al. [5], in addition to strategic responses, refugees' responses may be affected by feelings of insecurity, lack of trust, or insufficient understanding of how the information will be used. Inaccurately capturing refugees' data and preferences can ultimately result in flawed matches.

Matching Tools. GeoMatch relies on administrative data collected by the resettlement agency Global Refuge, and data from the ZEMIS database in Switzerland. Annie™ Moore relies on administrative data collected by HIAS. Historical bias is a plausible concern in all three data sources, yet it is neither explicitly acknowledged nor addressed in the proposed matching tools. Further, representation bias may be present in all three data sources. While the data from the two U.S. agencies includes all resettled refugees, there may be selection bias: agencies use specific criteria to decide which refugees to resettle – and therefore which data to collect [99]. Similarly, in the

Switzerland case, although the ZEMIS database provides comprehensive records of all asylum seekers, the data used in GeoMatch is restricted to asylum seekers granted an F permit within a five-year window [14]. If certain refugee groups were more likely to receive this protection status during that period, the data set may exhibit representation bias, as not all asylum seekers would have an equal chance of being included.

Match'In and Re:Match rely on survey data, which makes them prone to representation bias and measurement error. The Match'In tool uses surveys to collect information from two groups: host municipalities and refugees accommodated in initial reception centers. Given that municipalities were not selected randomly but based on suitability and feasibility criteria for the project, participating municipalities may not present the best matches across the full spectrum of refugees' needs and preferences [100]. Additionally, self-selection bias may be introduced as refugees and municipalities decide whether to participate.

Further, measurement error can arise from the questionnaire design. Since refugees know the potential outcomes of the process (i.e., the municipalities participating in the project), they may tailor their responses to strategically influence the matching results, causing deviations from their true preferences. The factors cited by Ahani et al. [5] as reasons for such deviations similarly apply to Match'In. The Re:Match project shares similarities with Match'In in its data collection methodology, subjecting it to similar biases.

3.2 Target Variable

Overview. The target variable is a further critical component of algorithmic systems and plays a central role in fairness discussions [63, 73, 126]. The target variable of interest is often not directly observable. To this end, researchers usually use one or multiple proxy variables derived from observed data to operationalize the unobservable target [63]. Prominent examples include using criminal history or re-arrests as a proxy for recidivism risk [54, 73], or health care costs as a proxy for a person's health needs [92]. The target variable may be subject to several biases [63]: outcome measurement error, selective labels, selection bias, and confounding bias. The first bias arises when there is a discrepancy between the proxy and the true target. To mitigate such mismatch, measurement models with multiple proxy variables may be employed [28, 73]. Selective labels occur when the proxy (target) is only observed for groups that were exposed to a historical decision or treatment. In such cases, the counterfactual, how the outcome would have differed under an alternative decision or treatment, remains unknown [42]. Several methods are proposed to evaluate model performance in the presence of selective labels [38, 81]. Further, the selection bias discussed in Section 3.1 may exacerbate the problems of selective labels. This bias occurs if independent variables, as well as unobservables, affect the decision or treatment. Moreover, if these variables additionally affect the outcome, confounding bias is introduced [96].

Refugee-Location Matching Context. The matching tools are developed to improve refugee integration, making this the primary target variable. However, integration represents a multidimensional concept whose definition is highly debated among scholars [31]. The challenge lies in the multidimensional nature of integration as a process unfolding in multiple directions, spanning different dimensions, and involving different actors [31]. Integration may be conceptualized as a one-way (refugees adapt to the host society - assimilation) or two-way (both refugees and the host society adapt to each other) process [31]. Esser [49] views integration as a form of assimilation comprising four key dimensions: cognitive (adaption in knowledge, skills, and language proficiency), structural (adaption in rights, education, and employment), social (adaption in daily interactions), and emotional (sense of identification with host country). In contrast, Ager and Strang [4] understand integration as a two-way process, structured around four main dimensions defined as "markers and means": employment, housing, education, and health. Harder et al. [67] conceptualize integration across six dimensions: psychological, economic, political, social, linguistic, and navigational. The diversity of theoretical approaches highlights the complexity of conceptualizing the integration process and, therefore, the difficulty of measuring integration. To this end, matching tools typically rely on proxy variables. The most common proxies are indicators of economic integration, such as employment status [89]. However, since integration spans multiple dimensions (beyond economic integration), relying on a single proxy variable introduces a significant risk of outcome measurement error. Moreover, even when the proxy variable is meant to operationalize only one dimension of integration, measurement error may be present. For instance, next to employment status, other proxies such as job type, contract duration, and employment conditions are critical to measure economic integration [37].

While outcome measurement errors are a concern in both settings, prediction-based tools must also address selective labels, selection bias, and confounding bias. Using employment status as a proxy for integration can result in selective labels, as outcomes may only be observed for specific groups of refugees: in Germany, for instance, refugees whose employment bans are lifted early and those with shorter asylum processes are more likely to find employment due to reduced bureaucratic barriers [27, 29, 78]. As a result, positive employment outcomes may be underrepresented in data for refugees still facing bans or pending asylum procedures. Selection bias further exacerbates the issue of selective labels. For instance, studies have found that decisions such as asylum application approval are influenced by refugee characteristics like socioeconomic status and social capital [79]. This may lead to an over-representation of refugees with higher socioeconomic status among those employed in the data. Additionally, if location assignment is treated as a form of intervention, it is crucial to account for variables that affect both the assignment decision and the outcome to avoid confounding bias.

Matching Tools. GeoMatch and Annie™ Moore rely exclusively on labor market proxies to measure integration. Specifically, the proxies used are employment status 90 days after arrival in the U.S. context and employment status two, three, and four years after arrival in the Swiss context [5, 13]. Since only one proxy is used to operationalize the integration outcome, its validity is limited and subject to outcome measurement error. While this limitation is acknowledged in matching research, its consequences are not addressed [5, 13, 89]. Further, we point out the potential for confounding bias in the U.S. context. Refugees are not (quasi-)randomly assigned to local affiliates by resettlement agencies. Instead, various factors, including individual-level characteristics, are considered in the assignment process [70, 84]. This increases the risk of confounding bias, as many of these factors may be unobserved or otherwise not accounted for in model training but could have affected both the (historical) allocation decisions and the integration outcomes. For instance, mental health needs, which are considered by the resettlement agency Global Refuge in the resettlement decision, are not included in GeoMatch but can affect employment chances [41, 84]. In the use case of Switzerland, the (quasi-)random allocation of asylum seekers to cantons reduces the risk of confounding [111]. However, selection bias and selective labels may still be present. This can be the case when specific characteristics (e.g., country of origin) influence whether asylum seekers receive an F permit [39], or when inter-individual variations in the asylum process (e.g., the asylum process length) influence employment outcomes [65, 111].

The Match'In and Re:Match tools do not have a target variable in the sense of the label in supervised learning. However, the collected information on refugees' preferences for certain areas can be understood as proxy variable(s) for the target variable integration. Indeed, the tools aim to maximize these preferences by finding a municipality that best matches them. In this respect, Match'In and Re:Match, unlike the prediction-based tools, take into account multiple integration dimensions. They consider not only economic preferences but also preferences in areas like housing, language, and health [103, 112]. Additionally, Match'In allows refugees to double-weight one of the seven integration areas (see Table 1).

3.3 Features

Overview. The feature variables of matching systems must be carefully considered, not only because model accuracy and thus matching quality depend on them, but also because of their role in encoding social processes and shaping fairness outcomes [64]. Features may include (correlates of) protected attributes such as age, sex, race, and religious affiliation as defined in global anti-discrimination law [109]. At the very least, developers need to ensure that members of protected groups are not disadvantaged [50]. However, sensitive features may appear or emerge as combinations (e.g., gender and race), correlate with "non-sensitive" variables, or vary in definition and measurement across contexts [32]. Addressing the implications of these complexities is essential at all stages of the algorithm life cycle.

In addition, similar to the issues discussed in Section 3.2, fairness concerns may be rooted in the proxies used for the features, which in turn may suffer from measurement error [82]. A further role of features in connection with fairness is their contribution to Explainable Artificial Intelligence (XAI). In machine learning, a common objective of XAI is to balance prediction model complexity and interpretability [21, 85]. Including a large number of features with complex interactions may improve prediction accuracy at the expense of interpretability. This can complicate the identification and mitigation of bias sources [43] and reduce human trust and acceptance in these systems [108].

Refugee-Location Matching Context. Features of matching tools can be categorized into two levels: individual-level and location-level features. Individual-level features comprise refugee information, including socio-demographic characteristics, preferences, and needs. Location-level features include information about resettlement locations that may influence refugee integration, e.g., labour market opportunities, housing conditions, and medical and support services. Most of this information, like the target variable, must be operationalized into adequate variables. In addition, many individual-level features can be considered protected attributes according to global anti-discrimination law, particularly those related to racial and ethnic origin and religion. While there is no universal agreement on how and whether protected attributes should be included in predictive models, they are crucial for evaluating fairness outcomes [64].

In any case, careful attention must be paid to how these attributes are defined and measured. This is particularly the case for multi-dimensional constructs such as ethnicity, which may be measured in diverse ways, critically affecting the ability to identify adverse outcomes across different communities of refugees [75].

Matching Tools. According to the sensitive attribute categorization outlined by Simson et al. [109], both GeoMatch and Annie™ Moore have access to a range of protected attributes which are included as predictors in their prediction models[10] (thereof: sex, national origin, language, religion, family status, marital status, and age). At the same time, the total number of features included in both tools is limited. GeoMatch, in particular, includes eight individual-level predictors for the U.S. and ten for the Swiss context (see Table 1). Annie™ Moore incorporates eleven individual and two additional country-level features. Neither tool includes location-level features, despite research showing their importance in influencing refugee integration in the U.S. and Switzerland [90, 117]. Further, both tools incorporate feature interactions, both implicitly and explicitly. GeoMatch draws on gradient-boosted trees, limiting its direct interpretability. To facilitate transparency in matching decisions, however, these tools need to report the relationships identified between features and integration outcomes in each location. Annie™ Moore offers some transparency by providing a table of estimated coefficients for each location from their LASSO model. GeoMatch, however, does not report on the most relevant features and their relationship with the integration target.

In non-prediction-based tools, contextual individual and location-level information captured through survey questions can be interpreted as features (analogous to preferences representing the target variable). Both Match'In and Re:Match collect socio-demographic information about refugees, in addition to their preferences and needs. Some of these variables can again be understood as protected attributes. For both tools, however, it remains unclear whether biographical information (e.g., age) is directly used in the matching process. For Re:Match, the lack of publicly available information about the collected data makes it difficult to determine what individual-level information beyond preferences, including protected attributes, is used.

3.4 Concept Drift

Overview. Concept drifts broadly refer to different types of changes in the data distribution, which can significantly affect the tool's functioning [71]. Despite its importance, concept drifts are often

[10]For detailed information, see Table 3 in Appendix C.

overlooked during the design stage and are typically addressed in later downstream stages [22]. We argue that these considerations should be integrated into the design phase of matching tools: proactively accounting for concept drift beforehand ensures that the tool's architecture is better equipped to anticipate and mitigate its potential impacts [53]. In supervised learning, a mismatch between the training and the deployment data can lead to concept drift [62]. Bayram et al. [22] provide a comprehensive overview of various types of concept drift, of which two are particularly relevant in this context [22]: (1) shifts that occur when the distribution of the input variables changes, and (2) shifts that occur when the joint distribution of input and target variables changes, either with or without changes in the marginal distribution of the input data. Outside of prediction settings, drifts may cause a mismatch between the data that was collected and the true information it is sought to represent, for instance, due to changes in local conditions or preferences over time [22, 35, 102]. In case of survey data, two main considerations should be kept in mind. First, the frequency of survey updates, and second, the extent to which the questions, especially those capturing preferences, remain generalizable over time.

Refugee-Location Matching Context. A shift in input data over time may occur due to a change in the population of refugees arriving in a host country. This can happen when individuals from previously unaffected countries or from countries not previously considered eligible for resettlement are in need of refuge. For instance, new conflicts such as the war in Ukraine and the conflict in Sudan have forced people to flee and apply for asylum in new host societies [121, 122]. Such abrupt and often unexpected events could even be understood as a defining characteristic of refugee movements and thus are critical to consider in the design of matching tools. Changes in resettlement eligibility criteria can also have notable consequences, as evidenced by the sharp increase in refugees admitted from Venezuela and Guatemala in the U.S [34, 46].

The second type of drift, which involves changes in the relationship between input and output over time, can occur when unexpected shocks alter, for example, labor market integration chances for refugees. These shocks may include adverse events that shift societal attitudes towards (groups of) refugees, such as incidents like the 9/11 attack [107], as well as further salient events [9, 51, 91].

In non-machine learning settings, drifts can occur when the location-level data used in the matching process becomes outdated. Since regional characteristics are highly dynamic and subject to economic and demographic changes, as well as affected by the ongoing resettlement process itself, maintaining accurate data is challenging. For example, when location data is only updated annually but refugees are assigned throughout the year, local resources like language course availability may fluctuate considerably over time, making the collected data less accurate.

Matching Tools. Studies extending and evaluating GeoMatch and Annie™ Moore raise concerns about distribution shifts [7, 16, 18]. However, these studies primarily focus on how shifts affect the matching stage rather than the prediction modeling stage. This focus arises because, to effectively match incoming refugees, the tools must determine the capacity of each location. Hence, the number of future arrivals informs the optimality criterion. To address this, Bansak and Paulson [18] propose for GeoMatch a dynamic matching algorithm that uses historical data to predict future arrivals, while Bansak et al. [16] introduce two additional algorithms that do not depend on historical data. For Annie™ Moore, Ahani et al. [7] emphasize the expertise of resettlement officers and their active monitoring of events that may affect the resettlement process. Therefore, Annie™ Moore allows officers to override predictions when alternative scenarios are expected. As a (rare) example of addressing shifts in the modeling stage, Bansak et al. [19] apply GeoMatch in the Netherlands and propose to account for shifts by predicting a long-term outcome alongside a correlated short-term proxy. Despite this extension and acknowledging the risk of drift, the tools do not address the potential impact of concept drift on (subgroup) model performance or provide details of any monitoring mechanisms. Furthermore, there is no information on how often the models are re-trained or how new training data affected by the tools' previous decisions is incorporated into the modeling process [53].

Turning to the non-prediction-based tools, Match'In, which was piloted for one year, did not update municipality information during the matching period. Noteworthy is, however, the following: the tool follows a Case-Based Reasoning (CBR) approach, which typically involves storing information about solved cases (here, the municipality to which the refugee was matched) in a knowledge base. This base is then used for matching newly arriving refugees [1, 103]. However, Match'In does not retain such information, which reduces the risk of the knowledge base becoming outdated, redundant, or inconsistent [103, 113]. Re:Match, in contrast, addresses potential drifts by updating municipality information before each cohort is matched [112]. Nonetheless, if these tools were implemented long-term, regular data updates and survey revisions would be necessary to keep the matching aligned with observed realities.

4 Empirical Illustration: Target Choice Sensitivity in GeoMatch

We empirically illustrate the importance of design decisions in algorithmic matching, focusing in particular on how different operationalizations of the target variable can influence matching outcomes. To this end, we simulate the application of the GeoMatch tool using German refugee survey data. Specifically, we model a hypothetical scenario in which asylum seekers arriving in Germany are assigned to one of Germany's 16 federal states by the GeoMatch tool. This contrasts with the standard resettlement procedure in Germany, which relies on the "Königsteiner Key"[12]. This key determines the quota of asylum seekers that each state must accommodate, calculated on the basis of two (weighted) factors: 2/3 tax revenue and 1/3 population. Once resettled, asylum seekers are restricted to reside in the assigned state for at least three years [10]. Paired with rich panel data on refugees, the (quasi-)random allocation makes Germany a strong case study, as it significantly reduces selection and confounding bias [116]. We use the case study to highlight the critical role of the target variable in location matching tools. In particular, we aim to demonstrate the implications of using different integration proxies.

4.1 Data Source, Target and Features

We use data from the IAB-BAMF-SOEP Survey of Refugees in Germany. The survey is conducted annually since 2016 and is integrated into the German Socio-Economic Panel (SOEP v38)[11] [30, 47]. Responsible for the study are the Institute for Employment Research (IAB), the Research Center of the Federal Office for Migration and Refugees (BAMF-FZ), and the SOEP. The survey collects representative information on refugees and asylum seekers who arrived in Germany since January 2013 by drawing random samples from the Central Register of Foreigners (AZR). The data provides detailed information on various characteristics and integration outcomes of refugees, making it a high-quality source for studies on refugee integration in Germany [8, 80, 116]. We use the survey to construct information for all asylum seekers aged 18-67 who arrived between 2013 and 2018, resulting in a total of 5,889 observations. Summary statistics can be found in Appendix D.

Target. We define two target variables. Each target captures one dimension of integration: economic and social integration [67]. This is the first application of GeoMatch using a target that goes beyond economic integration.

(1) Employment: This target measures whether asylum seekers found employment within three years of arrival in Germany (0 = "no", 1 = "yes"). We use a long-term outcome, similar to GeoMatch in Switzerland [13], since employment rates of refugees remain low in the first few years of arrival in Germany [29].
(2) Social activity: This target measures asylum seekers' social activity based on an assessment of their membership in organizations, frequency of social interactions, and participation in various activities (following Harder et al. [67]). The combined measure then determines whether an individual's social activity is above the median of others in the same survey year (0 = "no", 1 = "yes"). We provide an overview of the individual questions in Appendix D.

Features. We consider the following pre-arrival features as predictors for both targets: sex, age, immigration year, free-case, country of origin, religious affiliation, German level (speaking, writing, and reading), education level, school years, and vocational training. This is the same set of features included in the original GeoMatch applications [13], with two additions: school years and vocational training.

4.2 Analytical Approach

Our modeling approach includes four stages: data setup, modeling, mapping, and matching (following Bansak et al. [13]). In the first stage, we divide the data into train and test data. The train data consists of information on asylum seekers who arrived in Germany between 2013 and 2015, while the test data includes those who arrived between 2016 and 2018. We divide the train data into 16 subsets, one for each federal state. In the second stage, we fit a model on each train data subset using gradient-boosted trees [13, 57, 68]. Each model generates location-specific integration (proxy) predictions for each individual in the test set. For the target variable employment, the models predict the probability of a person finding employment within three years of arrival for each potential resettlement location. The same logic applies to the target variable social activity. In the third stage, we transform the individual-level predictions for each state to case-level predictions for individuals belonging to the same "case," such as family members. The case-level metric is the probability that at least one individual in the case finds employment/carries out a social activity in the respective state. In the final stage, cases in the test set are assigned to an "optimal" state. The optimal assignment is determined as the one that maximizes the respective global average integration probability. To solve the optimization problem, following Bansak et al. [13], a RELAX-IV cost flow solver is used. The optimization accounts for capacity constraints to ensure that no state receives more cases than its defined capacity. The capacity is derived from the actual number of cases assigned to each state in the test data. This stage yields the following results for our evaluations: for each individual in the test set and both target variables, in addition to the actual (random) state assignment and observed outcome, we obtain the predicted outcome probabilities in all states and the respective algorithmic state assignment.

4.3 Results

We present the main results of our analysis in Figure 1. Panel A illustrates the gains achieved through algorithmic matching when GeoMatch maximizes economic integration. Under the observed (random) allocation process, the average employment rate is 15%. By matching refugees algorithmically, the rate increases to 31%, implying a relative gain of nearly 98%. The average social activity rate under the observed resettlement process is 28%. Social integration also improves under employment-optimized matching, with a relative increase of 14%. Results change considerably when we depart from the standard GeoMatch approach, which uses employment as a proxy for integration. Panel B presents the gains when social integration, instead of employment, is used as the target. In this case, the gain in social integration is significantly higher, with a relative increase of 56%, compared to the 14% observed in Panel A. Employment gains are lower when social integration is prioritized, but still achieve a 28% relative increase.

Changing the operationalization of the integration target leads to different location assignments for about 88% of refugees in the test set, further highlighting the susceptibility of algorithmic allocations to design decisions. In addition, we observe varying gains across groups (e.g., higher employment gains for women than for men), as well as an overall low prediction performance when applying the tool "out of the box" with our data (see Appendix D, Table 7 and 8).

5 Discussion

Location matching tools affect the lives of members of the most vulnerable communities – refugees that recently arrived in their new host country. While aiming to support their integration outcomes, algorithmic matching tools face significant risks to incorporate a range of errors and biases in model development and deployment due to the nuanced interplay between data, design decisions, and the multi-layered modeling steps common in location matching systems. Against this background, we provided a systematic comparison of the four major matching tools to date – GeoMatch, Annie™ Moore, Match'In, and Re:Match – point to weak spots and demonstrate the need for a holistic (re-)evaluation of common design choices in such systems given their considerable impacts.

[11] The refugee samples included in the SOEP are: M3, M4, M5, and M6 [76].

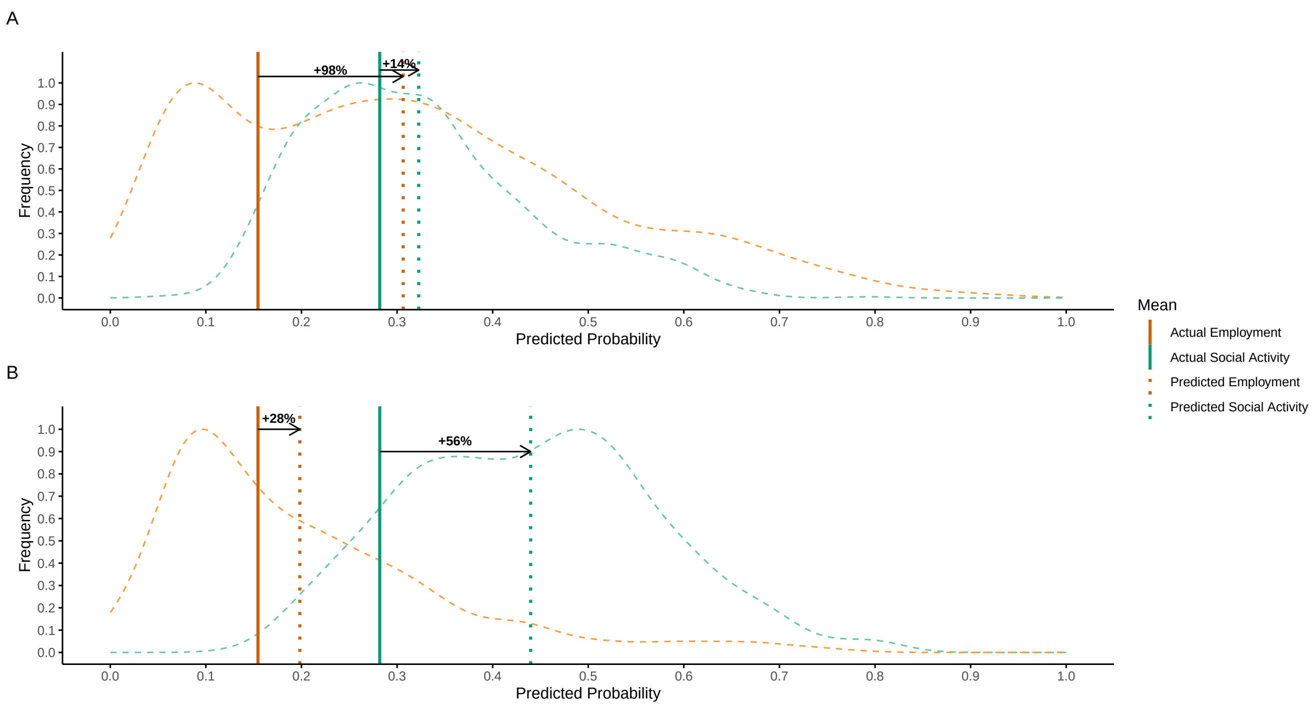

Figure 1: The prospective gains from algorithmic assignment heavily depend on the integration target optimized. The figure illustrates the prospective gains from assigning refugees algorithmically, compared to the observed resettlement process, when the objective was to maximize (A) employment and (B) social activity. For each scenario, distributions of the predicted probabilities for both targets are shown (dashed lines). The vertical lines indicate the actual (solid) and predicted mean employment/social activity rates (dotted).

In the first part of our research, we provided a comparative overview of the tools and their similarities and differences. This was prompted by the observation that existing research and the tools themselves often refer to each other without clearly stating the extent of their comparability [86, 89, 101]. The prediction-based tools (GeoMatch and Annie™ Moore) operate in two layers. The first layer involves a binary classification problem where machine learning is used for predictions, and the second layer utilizes these predictions in the matching process. Non-prediction-based tools (Match'In and Re:Match) directly perform the matching using similarity measures.

We highlight that, with the exception of Re:Match, none of the tools have published evaluation results of their pilot studies, contributing to an opaque view of their effectiveness and making it challenging to assess their impact.

The algorithmic fairness community has consistently emphasized the role of data, target, features, and deployment dynamics in shaping the social impacts of algorithmic models. In the second part of our study, we mapped this knowledge to allocation algorithms. Our mapping underscores that tools relying on machine learning have only been partially assessed with respect to the fairness implications of their design components. Open challenges include addressing historical biases in the training data, the operationalization and selection of the integration target and features, and the capacity to handle concept drift. These tools have instead primarily focused on the second layer, the matching stage [7, 17, 18, 56]. In contrast, the non-prediction-based tools have addressed specific limitations of prediction-based tools through a different design approach. They consider multiple integration dimensions, refugees' preferences, and the importance of data updates to account for shifts (notably Re:Match). However, other risks of biases remain unaddressed, including selection bias, measurement error, and the (ambiguous) role of protected attributes in matching.

In the last part, we exemplified the critical role of carefully conceptualizing the target for the allocation algorithms. We did this by applying GeoMatch to data from Germany and optimizing for social integration next to the common employment outcome. Our results show that the choice of the target has far-reaching implications for the matching process and the resulting integration gains. We find that social integration gains are much lower when the tool maximizes employment (and vice versa), and about 88% of refugees would be sent to a different location when modifying the integration target. We underline that these results do not advocate prioritizing one integration outcome over another but rather demonstrate the susceptibility of matching algorithms and the need for a more comprehensive understanding of multidimensional concepts such as integration in the matching context.

We acknowledge that our case study relies on survey data rather than official administrative data, which would likely be used in an actual implementation of GeoMatch in the German context. Additionally, we implemented GeoMatch "out-of-the-box" with basic state-level capacity constraints. Finally, the two proxy targets chosen to measure integration outcomes represent two specific design options, and further research is needed to understand the impact of other design decisions on matching outcomes across diverse groups of refugees.

We strongly encourage the algorithmic fairness community to apply their knowledge in the context of location matching. We suggest, in particular for the prediction-based tools, the development of new evaluation, auditing, and bias mitigation techniques that consider both interconnected layers of these tools. We further emphasize the need for systematic evaluation strategies for non-prediction-based matching tools, as the literature on fairness for these systems is particularly scarce.

6 Ethical Considerations Statement

In the course of our research, we encountered and addressed three main ethical challenges.

(1) In our main section, we reviewed the fairness implication of the design components of the four matching tools. We are aware that our comparison may be perceived as either criticizing or favoring certain tools. However, we deliberately refrained from making judgmental comments and rather aimed at identifying gaps in current location matching practice and pointing out instances where certain analyses or tests are currently missing. We believe it is crucial for the research groups responsible for these tools to acknowledge current limitations and address potential biases, given the tools' direct impact on people's lives and their implementation in multiple countries.

(2) For our use case, we constructed two proxies to measure two dimensions of integration. We fully recognize that these proxies represent only one of many possible approaches to assessing these dimensions. We also acknowledge that integration is a complex concept encompassing multiple dimensions beyond the two we have focused on in our research.

(3) We used individual-level data of asylum seekers in Germany in our empirical analysis. To ensure the privacy of the respondents, we took the following measures: 1) we followed the official procedure to obtain data from the SOEP and signed a nondisclosure agreement. 2) We agreed to keep the data confidential and to use it only for the intended research project. 3) We did not make any material publicly available that could disclose the obtained data.

7 Adverse Impacts Statements

Our research may have two crucial unintended impacts.

(1) The comparison of tools and the enumeration of potential biases for each tool may be used to favor one tool over another. However, the number of challenges we list for each tool should not be taken as a direct indication of its quality. Our purpose in listing sources of biases is to raise awareness of potential problems in different contexts, and great care is needed in weighing risks and potential benefits of implementing a particular tool in practice.

(2) The results of our use case may be subject to misinterpretation, specifically the findings indicating substantial gains in employment and social activity when asylum seekers are assigned algorithmically. To prevent misuse, we stress two key points. First, these results do not come from a full-scale evaluation of an implemented project but from a hypothetical test based on survey data with a specific evaluation focus. Therefore, we do not imply that algorithmic allocation necessarily improves over the current allocation process in Germany across all relevant measures and dimensions. Second, while the algorithmic assignment shows higher overall gains when using the employment compared to the social activity target, this should not be interpreted as a preference for optimizing employment over social activity. On the contrary, we argue that matching tools should recognize that integration is a multi-dimensional process.

Acknowledgments

The work was conducted as part of the grant "Consequences of Artificial Intelligence for Urban Societies (CAIUS), Germany ", funded by Volkswagen Foundation. Further, the work is supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

References

[1] Agnar Aamodt and Enric Plaza. 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications* 7, 1 (1994), 39–59. https://doi.org/10.3233/AIC-1994-7104

[2] Avidit Acharya, Kirk Bansak, and Jens Hainmueller. 2022. Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. *Political Analysis* 30, 1 (Jan. 2022), 89–112. https://doi.org/10.1017/pan.2020.48

[3] Avidit Acharya, Kirk Bansak, and Jens Hainmueller. 2022. Replication Data for: Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. https://doi.org/10.7910/DVN/ZEV0WX

[4] Alastair Ager and Alison Strang. 2008. Understanding Integration: A Conceptual Framework. *Journal of Refugee Studies* 21, 2 (June 2008), 166–191. https://doi.org/10.1093/jrs/fen016

[5] Narges Ahani, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C. Trapp. 2021. Placement Optimization in Refugee Resettlement. *Operations Research* 69, 5 (Sept. 2021), 1468–1486. https://doi.org/10.1287/opre.2020.2093

[6] Narges Ahani, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C. Trapp. 2021. Supplementary Material: Placement Optimization in Refugee Resettlement. *Operations Research* 69, 5 (Sept. 2021), 1468–1486. https://doi.org/10.1287/opre.2020.2093

[7] Narges Ahani, Paul Gölz, Ariel D. Procaccia, Alexander Teytelboym, and Andrew C. Trapp. 2024. Dynamic Placement in Refugee Resettlement. *Commun. ACM* 67, 5 (May 2024), 99–106. https://doi.org/10.1145/3611073

[8] Cevat Giray Aksoy, Panu Poutvaara, and Felicitas Schikora. 2023. First Time around: Local Conditions and Multi-Dimensional Integration of Refugees. *Journal of Urban Economics* 137 (Sept. 2023), 103588. https://doi.org/10.1016/j.jue.2023.103588

[9] Dimitris C Anagnostopoulos, George Giannakopoulos, and Nikos G Christodoulou. 2017. The Synergy of the Refugee Crisis and the Financial Crisis in Greece: Impact on Mental Health. *International Journal of Social Psychiatry* 63, 4 (June 2017), 352–358. https://doi.org/10.1177/0020764017700444

[10] Ludger Baba, Marco Schmandt, Constantin Tielkes, and Felix Weinhardt. 2024. *Evaluation der Wohnsitzregelung nach § 12a AufenthG.* Technical Report 13. Bundesamt für Migration und Flüchtlinge. https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Beitragsreihe/beitrag-band-13-evaluation-wohnsitzregelung.html?nn=283560

[11] BAMF. 2024. *Germany4ukraine.* Retrieved 2025-01-13 from https://www.germany4ukraine.de/hilfeportal-en
[12] BAMF. 2024. *Initial Distribution of Asylum Seekers (EASY).* Retrieved 2024-11-05 from https://www.BAMF.de/EN/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilung-node.html
[13] Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. Improving Refugee Integration through Data-Driven Algorithmic Assignment. *Science* 359, 6373 (Jan. 2018), 325–329. https://doi.org/10.1126/science.aao4408
[14] Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. Supplementary Material: Improving Refugee Integration through Data-Driven Algorithmic Assignment. *Science* 359, 6373 (Jan. 2018), 325–329. https://doi.org/10.1126/science.aao4408
[15] Kirk Bansak and Jens Hainmueller. 2017. Replication Code for: Improving Refugee Integration Through Data-Driven Algorithmic Assignment. https://doi.org/10.7910/DVN/MS8XES
[16] Kirk Bansak, Soonbong Lee, Vahideh Manshadi, Rad Niazadeh, and Elisabeth Paulson. 2024. Dynamic Matching with Post-allocation Service and Its Application to Refugee Resettlement. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.4748762 social science research network:4748762
[17] Kirk Bansak and Linna Marten. [n. d.]. Algorithmic Decision-Making, Fairness, and the Distribution of Impact: Application to Refugee Matching in Sweden. ([n. d.]).
[18] Kirk Bansak and Elisabeth Paulson. 2024. Outcome-Driven Dynamic Refugee Assignment with Allocation Balancing. *Operations Research* (March 2024). https://doi.org/10.1287/opre.2022.0445
[19] Kirk C. Bansak, Elisabeth Paulson, and Dominik Rothenhaeusler. 2024. Learning Under Random Distributional Shifts. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 3943–3951.
[20] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2023. Fairness and Machine Learning: Limitations and Opportunities. *MIT Press* (2023).
[21] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* 58 (June 2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
[22] Firas Bayram, Bestoun S. Ahmed, and Andreas Kassler. 2022. From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors. *Knowledge-Based Systems* 245 (June 2022), 108632. https://doi.org/10.1016/j.knosys.2022.108632
[23] Lucía Salgado Beirens, Hanne. 2023. *What Role Could Digital Technologies Play in the New EU Pact on Migration and Asylum?* Technical Report. Migration Policy Institute.
[24] Petra Bendel, Katharina Euler, Danielle Kasparick, Felix Maas, Sonja Reinhold, and Hannes Schammann. 2024. Match'In Questionnaires for People Seeking Protection and Municipalities. (April 2024). https://doi.org/10.25528/191
[25] Ralph Bergmann, Klaus-Dieter Althoff, Mirjam Minor, Meike Reichle, and Kerstin Bach. 2009. Case-Based Reasoning - Introduction and Recent Developments. *Künstliche Intelligenz* 1/2009 (Jan. 2009), 5–11.
[26] Paul P. Biemer. 2010. Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly* 74, 5 (Jan. 2010), 817–848. https://doi.org/10.1093/poq/nfq058
[27] BMAS. 2024. *BMAS - Arbeitsmarktzugang für Geflüchtete.* Retrieved 2024-12-26 from https://www.bmas.de/DE/Arbeit/Migration-und-Arbeit/Flucht-und-Aysl/Arbeitsmarktzugang-fuer-Gefluechtete/arbeitsmarktzugang-fuer-gefluechtete.html
[28] Laura Boeschoten, Erik-Jan van Kesteren, Ayoub Bagheri, and Daniel L. Oberski. 2021. Achieving Fair Inference Using Error-Prone Outcomes. *International Journal of Interactive Multimedia and Artificial Intelligence* 6, Special Issue on Artificial Intelligence, Paving the Way to the Future (2021), 9–15.
[29] Herbert Brücker. 2024. Arbeitsmarktintegration von Geflüchteten: Verbesserte institutionelle Rahmenbedingungen fördern die Erwerbstätigkeit. *IAB* (2024).
[30] Herbert Brücker, Nina Rother Rother, and Jürgen Schupp. 2017. IAB-BAMF-SOEP Befragung von Geflüchteten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. *IAB Forschungsbericht* 13 (2017). https://doi.org/10.5684/soep.core.v38eu
[31] Stephen Castles, Maja Korac, Ellie Vasta, and Steven Vertovec. 2003. *Integration: Mapping the Field.* Home Office. Research, Development and Statistics Directorate.
[32] Simon Caton and Christian Haas. 2024. Fairness in Machine Learning: A Survey. *ACM Comput. Surv.* 56, 7 (April 2024), 166:1–166:38. https://doi.org/10.1145/3616865
[33] Laura Celeste, Elisa Ertl, Craig Damian Smith, and Katja Wagner. 2024. *From Arrival to Belonging: Six Months Results of Re:Match.* Technical Report.
[34] Refugee Processing Center. 2024. *Refugee Arrivals by Nationality and Religion Group and Religion.* Retrieved 2024-11-26 from https://www.wrapsnet.org/archives/
[35] Peter Christen. 2012. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31164-2
[36] Buket Cilali, Kash Barker, and Andrés D. González. 2021. A Location Optimization Approach to Refugee Resettlement Decision-Making. *Sustainable Cities and Society* 74 (Nov. 2021), 103153. https://doi.org/10.1016/j.scs.2021.103153
[37] European Commission. 2024. *Labour Market Integration.* Retrieved 2024-10-25 from https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/labour-market-integration_en
[38] Amanda Coston, Alan Mishler, Edward H. Kennedy, and Alexandra Chouldechova. 2020. Counterfactual Risk Assessments, Evaluation, and Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20).* Association for Computing Machinery, New York, NY, USA, 582–593. https://doi.org/10.1145/3351095.3372851
[39] Swiss Refugee Council. 2024. *Residence Status.* Retrieved 2025-01-17 from https://www.refugeecouncil.ch/topics/asylum-in-switzerland/residence-status
[40] Swiss Refugee Council. 2024. *Short Overview of the Reception System.* Retrieved 2024-12-12 from https://asylumineurope.org/reports/country/switzerland/reception-conditions/short-overview-of-the-reception-system-2/
[41] Hai-Anh H. Dang, Trong-Anh Trinh, and Paolo Verme. 2023. Do Refugees with Better Mental Health Better Integrate? Evidence from the Building a New Life in Australia Longitudinal Survey. *Health Economics* 32, 12 (Dec. 2023), 2819–2835. https://doi.org/10.1002/hec.4750
[42] Maria De-Arteaga, Artur Dubrawski, and Alexandra Chouldechova. 2018. Learning under Selective Labels in the Presence of Expert Consistency. arXiv:1807.00905
[43] Luca Deck, Jakob Schoeffer, Maria De-Arteaga, and Niklas Kühl. 2024. A Critical Survey on Fairness Benefits of Explainable AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24).* Association for Computing Machinery, New York, NY, USA, 1579–1595. https://doi.org/10.1145/3630106.3658990
[44] David Delacretaz, Scott Kominers, and Alexander Teytelboym. 2024. Data and Code for: Matching Mechanisms for Refugee Resettlement. https://doi.org/10.3886/E191062V1
[45] David Delacrétaz, Scott Duke Kominers, and Alexander Teytelboym. 2023. Matching Mechanisms for Refugee Resettlement. *American Economic Review* 113, 10 (Oct. 2023), 2689–2717. https://doi.org/10.1257/aer.20210096
[46] DHS. 2023. *Refugees and Asylees Annual Flow Reports.* Retrieved 2024-11-26 from https://ohss.dhs.gov/topics/immigration/refugees-and-asylees/refugees-and-asylees-annual-flow-report
[47] DIW. 2023. IAB-BAMF-SOEP-Befragung Geflüchteter 2021, Daten der Jahre 2016-2021. https://doi.org/10.5684/soep.iab-bamf-soep-mig.2021
[48] Elisa Ertl, Laura Celeste, Giulia Fellin, and Lea Rau. 2024. *Data-Driven Matching of Refugees and Cities: An Implementation Guide.* Technical Report.
[49] Hartmut Esser. 2000. Assimilation, Integration und ethnische Konflikte. In *Migranten und Medien: Neue Herausforderungen an die Integrationsfunktion von Presse und Rundfunk*, Heribert Schatz, Christina Holtz-Bacha, and Jörg-Uwe Nieland (Eds.). VS Verlag für Sozialwissenschaften, Wiesbaden, 25–37. https://doi.org/10.1007/978-3-663-07794-7_2
[50] Alessandro Fabris, Andrea Esuli, Alejandro Moreo, and Fabrizio Sebastiani. 2023. Measuring Fairness Under Unawareness of Sensitive Attributes: A Quantification-Based Approach. *Journal of Artificial Intelligence Research* 76 (April 2023), 1117–1180. https://doi.org/10.1613/jair.1.14033
[51] Mariella Falkenhain, Uwe Flick, Andreas Hirseland, Shahed Naji, Kristina Seidelsohn, and Thomas Verlage. 2021. Setback in Labour Market Integration Due to the Covid-19 Crisis? An Explorative Insight on Forced Migrants' Vulnerability in Germany. *European Societies* 23, sup1 (Feb. 2021), S448–S463. https://doi.org/10.1080/14616696.2020.1828976
[52] Jeremy Ferwerda, Nicholas Adams-Cohen, Kirk Bansak, Jennifer Fei, Duncan Lawrence, Jeremy M. Weinstein, and Jens Hainmueller. 2020. Leveraging the Power of Place: A Data-Driven Decision Helper to Improve the Location Decisions of Economic Immigrants. https://doi.org/10.48550/arXiv.2007.13902
[53] Unai Fischer-Abaigar, Christoph Kern, Noam Barda, and Frauke Kreuter. 2024. Bridging the Gap: Towards an Expanded Toolkit for AI-driven Decision-Making in the Public Sector. *Government Information Quarterly* 41, 4 (Dec. 2024), 101976. https://doi.org/10.1016/j.giq.2024.101976
[54] Riccardo Fogliato, Alice Xiang, Zachary Lipton, Daniel Nagin, and Alexandra Chouldechova. 2021. On the Validity of Arrest as a Proxy for Offense: Race and the Likelihood of Arrest for Violent Crimes. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (July 2021), 100–111. https://doi.org/10.1145/3461702.3462538
[55] Flavia Fossati, Carlo Knotz, Fabienne Liechti, and Ihssane Otmani. 2024. The Gender Employment Gap among Refugees and the Role of Employer Discrimination: Experimental Evidence from the German, Swedish and Austrian

Labor Markets. *International Migration Review* 58, 1 (March 2024), 147–172. https://doi.org/10.1177/01979183221134274

[56] Daniel Freund, Thodoris Lykouris, Elisabeth Paulson, Bradley Sturt, and Wentao Weng. 2024. Group Fairness in Dynamic Refugee Assignment. https://doi.org/10.48550/arXiv.2301.10642

[57] Jerome H. Friedman. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 5 (2001), 1189–1232. jstor:2699986

[58] Frederic Gerdon, Ruben L Bach, Christoph Kern, and Frauke Kreuter. 2022. Social impacts of algorithmic decision-making: A research agenda for the social sciences. *Big Data & Society* 9, 1 (2022). https://doi.org/10.1177/20539517221089305

[59] Global Compact on Refugees. 2024. *GeoMatch: Connecting People to Places Using Artificial Intelligence.* Retrieved 2024-11-05 from http://globalcompactrefugees.org/good-practices/geomatch-connecting-people-places-using-artificial-intelligence

[60] Sylvie Graf, Mark Rubin, Yvette Assilamehou-Kunz, Mauro Bianchi, Andrea Carnaghi, Fabio Fasoli, Eerika Finell, Marie Gustafsson Sendén, Soraya Elizabeth Shamloo, Jaroslav Tocik, David Lacko, and Sabine Sczesny. 2023. Migrants, Asylum Seekers, and Refugees: Different Labels for Immigrants Influence Attitudes through Perceived Benefits in Nine Countries. *European Journal of Social Psychology* 53, 5 (2023), 970–983. https://doi.org/10.1002/ejsp.2947

[61] Robert M. Groves and Lars Lyberg. 2010. Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* 74, 5 (Jan. 2010), 849–879. https://doi.org/10.1093/poq/nfq065

[62] Cornelia Gruber, Patrick Oliver Schenk, Malte Schierholz, Frauke Kreuter, and Göran Kauermann. 2023. Sources of Uncertainty in Machine Learning – A Statisticians' View. https://doi.org/10.48550/arXiv.2305.16703 arXiv:2305.16703

[63] Luke Guerdan, Amanda Coston, Zhiwei Steven Wu, and Kenneth Holstein. 2023. Ground(Less) Truth: A Causal Framework for Proxy Labels in Human-Algorithm Decision-Making. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 688–704. https://doi.org/10.1145/3593013.3594036

[64] Maryam Amir Haeri and Katharina Anna Zweig. 2020. The Crucial Role of Sensitive Attributes in Fair Classification. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2993–3002. https://doi.org/10.1109/SSCI47803.2020.9308585

[65] Jens Hainmueller, Dominik Hangartner, and Duncan Lawrence. 2016. When Lives Are Put on Hold: Lengthy Asylum Processes Decrease Employment among Refugees. *Science Advances* 2, 8 (Aug. 2016), e1600432. https://doi.org/10.1126/sciadv.1600432

[66] Cyrine Hannafi and Mohamed Ali Marouani. 2023. Social Integration of Syrian Refugees and Their Intention to Stay in Germany. *Journal of Population Economics* 36, 2 (April 2023), 581–607. https://doi.org/10.1007/s00148-022-00913-1

[67] Niklas Harder, Lucila Figueroa, Rachel M. Gillum, Dominik Hangartner, David D. Laitin, and Jens Hainmueller. 2018. Multidimensional Measure of Immigrant Integration. *Proceedings of the National Academy of Sciences of the United States of America* 115, 45 (Nov. 2018), 11483–11488. https://doi.org/10.1073/pnas.1808793115

[68] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

[69] Sonja Haug, Susanne Lochner, and Dominik Huber. 2019. Methodological Aspects of a Quantitative and Qualitative Survey of Asylum Seekers in Germany – A Field Report. *methods, data, analyses* 13, 2 (July 2019), 20. https://doi.org/10.12758/mda.2019.02

[70] HIAS. [n. d.]. *Resettle Refugees.* Retrieved 2025-01-10 from https://hias.org/what/resettle-refugees/

[71] Fabian Hinder, Valerie Vaquet, Johannes Brinkrolf, and Barbara Hammer. 2023. Model-Based Explanations of Concept Drift. *Neurocomputing* 555 (Oct. 2023), 126640. https://doi.org/10.1016/j.neucom.2023.126640

[72] Meltem Ineli-Ciger. 2023. Artificial Intelligence and Resettlement of Refugees: Implications for the Fundamental Rights. https://doi.org/10.2139/ssrn.4530427 social science research network:4530427

[73] Abigail Z. Jacobs and Hanna Wallach. 2021. Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 375–385. https://doi.org/10.1145/3442188.3445901

[74] Jannes Jacobsen. 2020. *Establishing a Survey of Refugees in Germany: Challenges in Sampling, Field Work and Measurement.* Ph. D. Dissertation. Kultur-, Sozial- und Bildungswissenschaftliche Fakultät. https://doi.org/10.18452/21856

[75] Sofia Jaime and Christoph Kern. 2024. Ethnic Classifications in Algorithmic Fairness: Concepts, Measures and Implications in Practice. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 237–253. https://doi.org/10.1145/3630106.3658902

[76] Selin Kara and Stefan Zimmermann. 2023. *SOEPcompanion (V38)*. Research Report 1261. SOEP Survey Papers.

[77] Steve Kirkwood, Simon Goodman, Chris McVittie, and Andy McKinlay. 2016. Who Counts as an Asylum-Seeker or Refugee? In *The Language of Asylum: Refugees and Discourse*, Steve Kirkwood, Simon Goodman, Chris McVittie, and Andy McKinlay (Eds.). Palgrave Macmillan UK, London, 78–95. https://doi.org/10.1007/978-1-137-46116-2_6

[78] Yuliya Kosyakova and Hanna Brenzel. 2020. The Role of Length of Asylum Procedure and Legal Status in the Labour Market Integration of Refugees in Germany. *Soziale Welt* 71, 1/2 (2020), 123–159. jstor:27004995

[79] Yuliya Kosyakova and Herbert Brücker. 2020. Seeking Asylum in Germany: Do Human and Social Capital Determine the Outcome of Asylum Procedures? *European Sociological Review* 36, 5 (Oct. 2020), 663–683. https://doi.org/10.1093/esr/jcaa013

[80] Yuliya Kosyakova and Marie-Christine Laible. 2024. Importance of Personality Traits for Destination-Language Acquisition: Evidence for Refugees in Germany. *International Migration Review* 58, 1 (March 2024), 347–385. https://doi.org/10.1177/01979183221132538

[81] Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Halifax NS Canada, 275–284. https://doi.org/10.1145/3097983.3098066

[82] Yiqiao Liao and Parinaz Naghizadeh. 2023. Social Bias Meets Data Bias: The Impacts of Labeling and Measurement Errors on Fairness Criteria. https://doi.org/10.48550/arXiv.2206.00137 arXiv:2206.00137

[83] Oliver Lipps and Georg Lutz. 2017. Gender of Interviewer Effects in a Multi-topic Centralized CATI Panel Survey. *Methods, Data, Analyses* 11, 1 (Feb. 2017), 20. https://doi.org/10.12758/mda.2016.009

[84] LIRS. 2021. *What Is the USA Refugee Resettlement Process?* Retrieved 2025-01-10 from https://www.globalrefuge.org/news/refugee-resettlement-process/

[85] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.

[86] Match'In. 2023. *Match'In – Pilotprojekt Zur Verteilung von Schutzsuchenden.* Retrieved 2025-01-15 from https://matchin-projekt.de/en/

[87] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2022. A Survey on Bias and Fairness in Machine Learning. https://doi.org/10.48550/arXiv.1908.09635 arXiv:1908.09635 [cs]

[88] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application* 8, Volume 8, 2021 (March 2021), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

[89] Ahmed Ezzeldin Mohamed and Craig Damian Smith. 2024. Ethically Informed Algorithmic Matching and Refugee Resettlement. https://doi.org/10.2139/ssrn.4848913

[90] Tobias Müller, Pia Pannatier, and Martina Viarengo. 2023. Labor Market Integration, Local Conditions and Inequalities: Evidence from Refugees in Switzerland. *World Development* 170 (Oct. 2023), 106288. https://doi.org/10.1016/j.worlddev.2023.106288

[91] Christof Nägel and Mark Lutter. 2020. The Christmas Market Attack in Berlin and Attitudes Toward Refugees: A Natural Experiment with Data from the European Social Survey. *European Journal for Security Research* 5, 2 (Oct. 2020), 199–221. https://doi.org/10.1007/s41125-020-00066-w

[92] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science (New York, N.Y.)* 366, 6464 (Oct. 2019), 447–453. https://doi.org/10.1126/science.aax2342

[93] OPSI. 2020. *Annie™ MOORE (Matching for Outcome Optimization and Refugee Empowerment) - Observatory of Public Sector Innovation.* Retrieved 2024-11-06 from https://oecd-opsi.org/innovations/annie/,https://oecd-opsi.org/innovations/annie/

[94] Overheid. 2024. *Promising Link with GeoMatch - Central Agency for the Reception of Asylum Seekers.* Retrieved 2024-11-05 from https://algoritmes.overheid.nl/en/algoritme/promising-link-with-geomatch-central-agency-for-the-reception-of-asylum-seekers/96961197#werking

[95] Derya Ozkul. 2023. Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe. (2023). https://doi.org/10.13140/RG.2.2.24295.46248

[96] Miquel Porta. 2016. A Dictionary of Epidemiology. In *A Dictionary of Epidemiology*. Oxford University Press.

[97] Jing Qin. 2017. *Biased Sampling, Over-identified Parameter Problems and Beyond.* Springer, Singapore. https://doi.org/10.1007/978-981-10-4856-2

[98] Lea Rau, Elisa Ertl, and Katja Wagner. 2024. *Re:Match as an Innovative Tool for the Relocation of Protection Seekers.* Technical Report.

[99] RCUSA. 2024. *Resettlement Process - Refugee Council USA.* Retrieved 2024-11-05 from https://rcusa.org/resources/resettlement-process/

[100] Sonja Reinhold and Katharina Euler. 2023. *Match'In – Pilotprojekt Zur Verteilung von Schutzsuchenden Mit Hilfe Eines Algorithmengestützten Matching-Verfahrens.*

[101] Re:Match. 2021. *Re:Match - Relocation of Refugees via Matching by Algorithm.* Retrieved 2025-01-15 from https://rematch-eu.org/

[102] Jing Ren, Feng Xia, Xiangtai Chen, Jiaying Liu, Mingliang Hou, Ahsan Shehzad, Nargiz Sultanova, and Xiangjie Kong. 2021. Matching Algorithms: Fundamentals, Applications and Challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence* 5, 3 (June 2021), 332–350. https://doi.org/10.1109/TETCI.2021.3067655
[103] Christian Sauer, Carsten Wenzel, Katharina Euler, Sonja Reinhold, Frank Wuttke, and Bjoern Oitmann. 2024. Match'In - Pilot Project of an Algorithm-Based Decision Support System for Individualized Recommendations of Municipalities for the Integration of Refugees. In *ECAI 2024*. IOS Press, 4703–4710. https://doi.org/10.3233/FAIA241067
[104] Katie Schwarzmann. 2024. Automated Recommendation-Making Tools in Immigration Systems. *The Churchill Fellowship* (2024).
[105] SEM. 2024. *Migrationsinformationssystem Der Schweiz Soll Erneuert Werden.* Retrieved 2024-12-09 from https://www.sem.admin.ch/sem/de/home/sem/medien/mm.msg-id-83179.html
[106] Nima Shahbazi, Yin Lin, Abolfazl Asudeh, and H. V. Jagadish. 2023. Representation Bias in Data: A Survey on Identification and Resolution Techniques. *ACM Comput. Surv.* 55, 13s (July 2023), 293:1–293:39. https://doi.org/10.1145/3588433
[107] Seonho Shin. 2021. Were They a Shock or an Opportunity?: The Heterogeneous Impacts of the 9/11 Attacks on Refugees as Job Seekers—a Nonlinear Multi-Level Approach. *Empirical Economics* 61, 5 (Nov. 2021), 2827–2864. https://doi.org/10.1007/s00181-020-01963-8
[108] Avital Shulner-Tal, Tsvi Kuflik, and Doron Kliger. 2023. Enhancing Fairness Perception – Towards Human-Centred AI and Personalized Explanations Understanding the Factors Influencing Laypeople's Fairness Perceptions of Algorithmic Decisions. *International Journal of Human–Computer Interaction* 39, 7 (April 2023), 1455–1482. https://doi.org/10.1080/10447318.2022.2095705
[109] Jan Simson, Alessandro Fabris, and Christoph Kern. 2024. Lazy Data Practices Harm Fairness Research. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 642–659. https://doi.org/10.1145/3630106.3658931
[110] Jan Simson, Florian Pfisterer, and Christoph Kern. 2024. One Model Many Scores: Using Multiverse Analysis to Prevent Fairness Hacking and Evaluate the Influence of Model Design Decisions. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1305–1320. https://doi.org/10.1145/3630106.3658974
[111] Michaela Slotwinski, Alois Stutzer, and Roman Uhlig. 2019. Are Asylum Seekers More Likely to Work with More Inclusive Labor Market Access Regulations? *Swiss Journal of Economics and Statistics* 155, 1 (Dec. 2019), 17. https://doi.org/10.1186/s41937-019-0046-2
[112] Craig Damian Smith, Laura Celeste, Ahmed Mohammed, Elisa Ertl, Katja Wagner, and Lea Rau. 2024. *Re:Match – Relocation via Matching.* Technical Report.
[113] Barry Smyth. 1998. Case-Base Maintenance. In *Tasks and Methods in Applied Artificial Intelligence*, Angel Pasqual del Pobil, José Mira, and Moonis Ali (Eds.). Springer, Berlin, Heidelberg, 507–516. https://doi.org/10.1007/3-540-64574-8_436
[114] Theoni Stathopoulou, Elica Krajčeva, Natalja Menold, and Steve Dept. 2019. Questionnaire Design and Translation for Refugee Populations: Lessons Learned from the REHEAL Study. *Journal of Refugee Studies* 32, Special_Issue_1 (Dec. 2019), i105–i121. https://doi.org/10.1093/jrs/fez045
[115] Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3465416.3483305
[116] Jasper Tjaden and Christoph Spörlein. 2023. How Much Do "Local Policies" Matter for Refugee Integration? An Analytical Model and Evidence from a Highly Decentralized Country. *International Migration Review* (Oct. 2023), 01979183231205561. https://doi.org/10.1177/01979183231205561
[117] Van C. Tran and Francisco Lara-García. 2020. A New Beginning: Early Refugee Integration in the United States. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 6, 3 (Nov. 2020), 117–149. https://doi.org/10.7758/RSF.2020.6.3.06
[118] UNHCR 2024. *Asylum-Seekers.* Retrieved 2024-11-11 from https://www.unhcr.org/about-unhcr/who-we-protect/asylum-seekers
[119] UNHCR. 2024. *Projected Global Resettlement Needs 2025.* Technical Report. The UN Refugee Agency.
[120] UNHCR 2024. *Refugees.* Retrieved 2024-11-11 from https://www.unhcr.org/about-unhcr/who-we-protect/refugees
[121] UNHCR. 2024. *Sudan Situation.* Retrieved 2024-12-02 from https://reporting.unhcr.org/operational/situations/sudan-situation
[122] UNHCR. 2024. *Ukraine Situation.* Retrieved 2024-12-27 from https://reporting.unhcr.org/operational/situations/ukraine-situation
[123] UNHCR. 2024. *UNHCR Viewpoint: 'Refugee' or 'Migrant' – Which Is Right?* Retrieved 2024-11-11 from https://www.unhcr.org/news/stories/unhcr-viewpoint-refugee-or-migrant-which-right
[124] U.S. Department of State. 2024. *Reception and Placement.* Retrieved 2024-11-20 from https://www.state.gov/refugee-admissions/reception-and-placement/
[125] Roos Van der Zwan and Frank Van Tubergen. 2024. Employment and Work Intentions of Refugee Women. Evidence on Syrians in the Netherlands. *Ethnic and Racial Studies* 0, 0 (2024), 1–22. https://doi.org/10.1080/01419870.2024.2388684
[126] Jamelle Watson-Daniels, Solon Barocas, Jake M. Hofman, and Alexandra Chouldechova. 2023. Multi-Target Multiplicity: Flexibility and Fairness in Target Specification under Resource Constraints. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. 297–311. https://doi.org/10.1145/3593013.3593998 arXiv:2306.13738 [cs]

A Glossary

Term	Definition	References
Asylum Seeker	An individual who has left their country of nationality and is seeking protection in another country, but whose refugee status or complementary protection status has not yet been processed. The legal definition is primarily based on the 1951 United Nations Convention On the Status of Refugees and its 1967 Protocol.	[77, 118]
Refugee	An individual who, according to the 1951 United Nations Convention On the Status of Refugees and its 1967 Protocol, has been forced to flee their country of nationality and cannot return due to a well-founded fear of persecution based on race, religion, nationality, social group, or political opinion. If the application for asylum has not yet been processed, the individual is defined as an asylum seeker.	[77, 120]
Migrant	An individual who chooses to leave their country of nationality not because of a direct threat of persecution or death, but for reasons such as work and education. A migrant does not apply for asylum, but for a visa or residence permit.	[60, 123]
Resettlement	A process that allows refugees to be resettled in another country with a status that guarantees international protection and, ultimately, permanent residency.	[119], p.9

Table 2: The table provides an overview of key terms and their definitions that lack a consistent and unambiguous definition in existing research [60].

B Background: Resettlement Process

B.1 U.S.

The resettlement process in the U.S., managed by the U.S. Refugee Admissions Program (USRAP), works as follows [5, 99, 124]. Initially, the United Nations High Commissioner for Refugees (UNHCR), U.S. embassies, or designated NGOs assess whether individuals seeking refugee status are eligible for resettlement in the U.S. These individuals undergo a rigorous screening process outside the U.S., which includes UNHCR assessments, interviews conducted by the Department of Homeland Security (DHS)/U.S. Citizenship and Immigration Services (USCIS), and medical and security checks. The president determines the number of admitted refugees to the U.S. each year. Once approved for resettlement, applicants are referred to one of the ten U.S. resettlement agencies. Two of these agencies are Global Refuge and HIAS. Each agency has a network of local affiliates across the country. The agencies decide to which local affiliate refugees will be assigned upon arrival. Placement decisions are made manually by agency staff, considering individual-level characteristics (such as language proficiency and country of origin) and location-specific factors (including housing availability, employment opportunities, and the capacity of local affiliates). Upon arrival, the agencies are responsible for providing housing and essential needs for the refugees. Additionally, refugees receive support through the Reception and Placement (R&P) program, which offers financial assistance and services for up to 90 days after arrival.

B.2 Switzerland

The resettlement process in Switzerland, managed by the State Secretariat for Migration (SEM), works as follows [13, 40, 111]. Upon arrival in Switzerland, asylum seekers are placed in one of the federal asylum centers within the six asylum regions run by the SEM. At these centers, it is determined whether an asylum request is to be approved or dismissed. If the request is not dismissed, asylum seekers are granted an N permit, indicating that their application is being processed and allowing them to reside in Switzerland during this period. Asylum seekers stay in the federal centers for a maximum of 140 days. If no decision is made within this period, or if the asylum seeker is assigned to the extended procedure, they are allocated to one of Switzerland's 26 cantons. The allocation decision is made according to Article 21(1) AO1, which states that each canton must receive a percentage of asylum seekers in proportion to its population. As no individual-level characteristics are included in the decision other than family reunification and special care needs, the decision can be considered (quasi-)random. After allocation, the cantons are responsible for providing accommodation and other services. Further, once assigned, asylum seekers cannot change their canton and must await the decision on their asylum request. The duration of the process can vary. If approved, they may receive a B permit, granting refugee status and a residence permit, or an F permit for temporary admission as refugees.

B.3 Germany

The resettlement process in Germany, managed by the Federal Office for Migration and Refugees (BAMF), works as follows [12, 66, 100]. Upon arrival in Germany, asylum seekers register at one of the nearest reception centers operated by BAMF. After registration, these reception centers use a computer-based system called EASY (Initial Distribution of Asylum Seekers) to assign asylum seekers to one of Germany's 16 federal states. The allocation decision is based on a distribution key called the "Königsteiner key." This key calculates the quota of asylum seekers each federal state must accommodate. The quota is re-calculated every year and is based on a weighted formula combining population and tax revenue. Specifically, 1/3 of the quota is determined by a state's share of the national population, and 2/3 by its share of national tax revenue. The allocation decision is considered (quasi-)random, as no factors other than the key and family reunification influence the assignment. Once assigned to a federal state, asylum seekers stay in the initial reception centers for a maximum of 18 months before being moved to municipalities within the state. The distribution to municipalities follows a similar quota-based system, though the specific allocation may vary depending on the federal state's policies. The asylum application is examined through document review and interviews throughout this process. Asylum seekers may receive approval for asylum, refugee protection status, subsidiary protection, or a deportation order.

The resettlement process for individuals fleeing Ukraine differs from the standard asylum procedure [11]. Individuals with Ukrainian nationality, their family members, and those with permanent residence in Ukraine do not need a visa or residence permit to enter Germany. Ukrainian refugees must register at one of the reception centers to receive government support upon arrival. Ukrainian refugees who have private accommodation or family relatives in Germany can register in those municipalities. Otherwise, a quota-based system called "FREE" assigns those needing accommodation to municipalities. Ukrainians must register and apply for a residence permit within 90 days of arrival.

C Matching Tools: Context and Methods

Formal Setup. We use the following notation to describe the tools. The number of refugees to be resettled upon arrival is indicated as $i \in \{1, ..., N\}$. Refugees are classified either as part of a "case" (usually a family unit) or as a "free-case". A refugee may be part of a case that includes members who have already been resettled. In this situation, the new arrival will generally be assigned to the same location as the previously resettled members. Similarly, all members will be assigned to the same location when a case arrives together. Refugee-location matching tools only consider refugees who are free-cases or cases arriving at the same time. The number of cases is denoted as $j \in \{1, ..., J\}$, with $J \leq N$. The locations to which refugees may be resettled upon arrival are indicated as $k \in \{1, ..., K\}$ (e.g., states, municipalities, or local affiliates of a resettlement agency). The time of arrival of refugees prior to resettlement can be specified in various formats, e.g. year, quarter, month, or day, and is denoted as t. Refugee-level characteristics are summarized as X_i, location-level characteristics as L_k. Each refugee's integration outcome (e.g., employment status) is observed only for the location to which they were actually assigned. Hence, if the refugee is assigned to location g, the observed outcome is $Y_i(g)$. According to the potential outcome framework, the outcome for the same refugee in a different location w, though unobserved, can be represented as $Y_i(w)$.

C.1 GeoMatch

Literature. The tool was originally introduced by Bansak et al. [13]. Since its introduction, several studies have evaluated and extended the tool's functionality. For instance, Ferwerda et al. [52] adapted it to serve as a decision support tool for economic migrants in Canada. Acharya et al. [2] added the option of incorporating refugee preferences into the matching stage using simulated preference data. Bansak and Marten [17] assessed the fairness of the tool using a potential outcomes approach with data from Sweden, while Freund et al. [56] explored group fairness in the matching stage. Additionally, Bansak et al. [19] presented various prediction methods to account for random distributional shifts using data from the Netherlands. Bansak et al. [16] introduced new matching algorithms that can account for distributional shifts. Bansak and Paulson [18] proposed an improved matching algorithm to ensure that the allocation of refugees across locations is balanced over time.

Method. The GeoMatch tool, as introduced by Bansak et al. [13], comprises four key stages: data setup, modeling, mapping, and matching. The integration outcome in GeoMatch is a binary variable indicating the employment status of refugees after a certain time period (for the U.S, 90 days after arrival, and for Switzerland, 2, 3, 4 years after arrival). In the first stage, historical data is split into train and test data according to a specific arrival time t. For instance, in the case of Switzerland, the train data includes all individuals who arrived between 1999 and 2012, while the test set consists of those who arrived in 2013. In the second stage, the train data is split into K subsets. For instance, one subset contains information on all individuals that were assigned to $k = 1$. A model is fitted to each of these subsets using various methods (gradient-boosted trees, random forests, elastic net regression, and kernel-based regularized least squares). Gradient-boosted trees are chosen as the primary prediction method due to their accuracy, calibration reliability, and variable selection capabilities. Each fitted model $\hat{f}_k(X)$ is used to predict the employment probability $\hat{Y}_i(k)$ for each i in the test set for each location k. Thus, for each i, K employment probability predictions $\hat{Y}_i(1), \hat{Y}_i(2), \ldots, \hat{Y}_i(K)$ are generated. However, only one true outcome is observed for each i. In the third stage, the individual-level employment probabilities are aggregated into case-level probabilities using various metrics. The primary metric is the probability that at least one individual in the case finds employment at a given location, computed as: $1 - \prod_{i \in j}(1 - \hat{Y}_i(k))$. Additional robustness metrics include the mean, maximum, and minimum probabilities of employment for each case. In the final stage, each case is assigned to the location that maximizes the global average employment probability while satisfying capacity constraints.

C.2 Annie™ Moore

Literature. The tool was originally introduced by Ahani et al. [5]. Delacrétaz et al. [45] added the option of incorporating refugee preferences into the matching stage using simulated preference data. Ahani et al. [7] released an extension, Annie™ Moore 2.0, that considers future refugee arrivals in the matching stage.

Method. Annie™ Moore, as introduced by Ahani et al. [5], follows four key stages similar to GeoMatch: data setup, modeling, mapping, and matching. The integration outcome in Annie™ Moore is a binary variable indicating the employment status of refugees 90 days after arrival. In the first stage, historical data is split into train and test sets according to a specific arrival time t. In the second stage, a LASSO logit model is fitted on the train data. The model includes interactions between resettlement locations and individual-level characteristics. Annie™ Moore, analogous to GeoMatch, also uses gradient-boosted trees for prediction. Both prediction methods perform similarly across various performance measures (misclassification error, recall, precision, and AUC-ROC). The LASSO model is selected as the primary prediction method and used to generate employment probability predictions for each individual i in the test data for each location k. In the third stage, the individual level probabilities are transformed to case-level probabilities. The case-level metric used is the sum of all individual predicted employment probabilities by case, computed as $\sum_{i \in j} \hat{Y}_i(k)$. Further robustness metrics are the mean, maximum, and minimum individual-level predicted employment probabilities for each case. In the final stage, each case is assigned to the location that maximizes an optimality criterion given existing constraints. The optimality criterion consists in maximizing the total expected number of employed refugees. Two types of constraints are included: 1) binary service constraints (language, nationality, single-parent, and large-family support) and 2) capacity constraints.

C.3 Match'In

Literature. The tool and methodology are presented by Sauer et al. [103].

Method. Match'In relies on Case-Base Reasoning (CBR), an experience based problem-solving methodology that simulates the human problem-solving approach [25]. The solving process, the CBR cycle, can be summarized in four steps according to Aamodt and Plaza [1]: 1. Retrieve, 2. Reuse, 3. Revise and 4. Retain. Meaning in order to solve a new case (new problem), similar cases from the case-base (collection of past problems) are *retrieved*, the solution of one or more similar cases *reused*, the constructed solution is *revised*, and the newly solved problem is *retained* into the knowledge-base [1, 25]. The Match'In tool operates in four main stages: feasibility evaluation, case generation, similarity calculation, and matching [100, 103]. In the first stage, the tool uses pre-defined exclusion criteria to determine whether an individual can be algorithmically matched. For instance, one such criterion is whether the individual has existing family ties in Germany. If this is the case, the individual will not be matched by the algorithm. In the second stage, if there are no exclusion criteria preventing a match, an ideal municipality is generated for the individual. This ideal municipality is created using a set of 64 pre-defined rules that translate the refugee's collected data into the characteristics of an optimal municipality. This ideal municipality serves as the new case in the CBR approach. In the third stage, the tool retrieves similar cases from the case base, which contains data on all participating municipalities. Similarity measures, defined in advance, are used to identify the municipality most closely matching the ideal case. Finally, before finalizing the match, a caseworker reviews and potentially revises the solution. Unlike standard CBR approaches, the revised solution is not stored in the knowledge base.

C.4 Re:Match

Literature. An overview of the first pilot phase in 2023 is provided by Smith et al. [112] and Celeste et al. [33]. The project further published a recommendation and an implementation guide [48, 98].

Method. The Re:Match tool operates in three key stages [48, 112]. In the first stage, a feasible set of municipalities is selected for each individual. Feasibility is determined by the size and composition of the individual's household and specific medical needs. In the second stage, a score is calculated for each feasible municipality by quantifying the match between refugees' background characteristics and preferences and municipalities' characteristics and capacities. Finally, municipalities are ranked according to their score. The top-ranked match is presented to the refugee, who has the option of accepting or rejecting the proposed location.

C.5 Protected Attributes

	GeoMatch [13, 14, 19]	**Annie™ [5, 6]**	**Match'In [24]**
Gender and Sexual Identity			
Sex	**!** (U.S., CH, NL)	**!** (U.S.)	
Sexual orientation			**!** (GER)
Gender			**!** (GER)
Racial and Ethnic Origin			
Race			
Color			
Ethnic origin	**!** (NL)		
National origin	Country of origin (U.S., CH, NL)	**!** (U.S.)	Passport (GER)
Language	English-speaking (U.S.), French-speaking (CH), Native language (NL)	**!** (U.S.)	**!** (GER)
National minority			
Socioeconomic Status			
Social origin			
Property			
Recipient of public assistance			
Religion, Belief and Opinion			
Religion	**!** (U.S., CH, NL)	**!** (U.S.)	**!** (GER)
Political opinion			
Other opinion			
Family			
Birth			
Familial status	**!** (U.S., CH)	**!** (U.S.)	**!** (GER)
Marital status	**!** (CH, NL)	**!** (U.S.)	**!** (GER)
Disability and Health Conditions			
Disability		Medical condition type (U.S.)	**!** (GER)
Genetic features			
Age			
Age	**!** (U.S., CH, NL)	**!** (U.S.)	**!** (GER)

Table 3: Matching tools draw on a number of protected attributes as defined by global anti-discrimination law. We follow the categorization presented by Simson et al. [109]. In line with the authors of this categorization, we indicate a ! if the feature of the tools matches the exact phrasing of the protected attribute in global anti-discrimination law; otherwise, we note the phrasing used. The abbreviations are as follows: U.S. for the United States, CH for Switzerland, and NL for the Netherlands. Information on the features, and consequently the protected attributes for GeoMatch in the Netherlands, is drawn from the study by Bansak et al. [19]. Further information on features for GeoMatch in the U.S. and Switzerland is drawn from Bansak et al. [14], for Annie™ Moore from Ahani et al. [6], and for Match'In from the survey questionnaire [24]. For Re:Match, no information on protected attributes is provided due to the lack of a publicly available questionnaire.

D Case Study: Additional Results

Target	% in Train Data	% in Test Data
Employment		
... Yes	21.38	15.44
... No	78.62	84.56
Social Activity		
... Yes	35.00	28.19
... No	65.00	71.81
Observations (N)	5112	777

Table 4: Summary statistics of integration targets.

Feature	Mean/% Train	Mean/% Test
Age at immigration	32.00	31.81
Immigration year	2014	2016
School years	9.9	9.38
Sex		
... Female	39.77 %	43.63 %
... Male	60.23 %	56.37 %
Free-case		
... No	46.13 %	34.62 %
... Yes	53.87 %	65.38 %
Country of origin		
... Syria	52.8 %	42.86 %
... Iraq	13.34 %	16.09 %
... Afghanistan	12.52 %	8.49 %
... Eritrea	4.32 %	7.34 %
... Iran	2.31 %	4.89 %
... Pakistan	1.72 %	1.16 %
... Russian Federation	1.45 %	0.51 %
... Somalia	1.19 %	4.38 %
... Kosovo	0.94 %	0 %
... Serbia	0.8 %	0.39 %
... Lebanon	0.63 %	0.51 %
... Albania	0.7 %	0.13 %
... Nigeria	0.68 %	3.99 %
... North Macedonia	0.35 %	0.13 %
... Ukraine	0.23 %	0.26 %
... Bosnia and Herzegovina	0.14 %	0 %
... Turkey	0.12 %	0.39 %
... India	0.08 %	0.39 %
... Romania	0.02 %	0 %
... Other	5.65 %	8.11 %
Religious affiliation		
... Catholic	2.51 %	2.87 %
... Protestant	2.65 %	3.78 %
... Christian Orthodox	1.03 %	2.35 %
... Other Christian religious denomination	6.65 %	4.69 %
... Islamic religious denomination	6.69 %	7.04 %
... Shiite religious denomination	60.64 %	56.45 %
... Sunnite religious denomination	7.61 %	10.43 %
... Alevite religious denomination	0.28 %	0 %
... Other religious denomination	6.17 %	6.52 %

Feature	Mean/% Train	Mean/% Test
... No religious denomination	5.78 %	5.87 %
Education level (cat: low)		
... No	66.11 %	61.29 %
... Yes	33.89 %	38.71 %
Education level (cat: medium)		
... No	73.22 %	72.16 %
... Yes	26.78 %	27.84 %
Education level (cat: high)		
... No	60.67 %	66.55 %
... Yes	39.33 %	33.45 %
German reading level		
... Very good	0.31 %	0.26 %
... Good	0.61 %	0.77 %
... Fairly	1.41 %	1.81 %
... Poorly	2.55 %	0.9 %
... None	95.12 %	96.26 %
German writing level		
... Very good	0.35 %	0.26 %
... Good	0.61 %	0.77 %
... Fairly	1.3 %	1.81 %
... Poorly	2.26 %	1.29 %
... None	95.49 %	95.87 %
German speaking level		
... Very good	0.73 %	1.29 %
... Good	1.12 %	1.81 %
... Fairly	1.37 %	1.94 %
... Poorly	2.85 %	1.55 %
... None	93.94 %	93.42 %
Vocational training		
... No	75.19 %	82.95 %
... Yes	24.81 %	17.05 %

Table 5: Summary statistics of features.

Category	Survey Question	Answer Options
Membership	Are you a member of a trade union?	Yes/No
	Are you a member of a professional body?	Yes/No
	Are you a member of a club or similar organization?	Yes/No
	Are you a member of a works or staff council?	Yes/No
Social Interaction	In the last 12 months, have ...	
	... you visited people of German origin in their home?	Yes/No
	... you visited people not from Germany or whose parents are not from Germany in their home?	Yes/No
	... people of German origin visited you in your home?	Yes/No
	... people not from Germany or whose parents are not from Germany visited you in your home?	Yes/No
Activity Frequency	How often do you ...	
	... visit neighbors or friends?	1 (Often) - 5 (Never)
	... visit family and relatives?	1 (Often) - 5 (Never)
	... attend sport events?	1 (Often) - 5 (Never)
	... spend time with German people?	1 (Often) - 5 (Never)
	... spend time with people from other countries?	1 (Often) - 5 (Never)
	Which of the following activities do you take part in during your free time ...	
	... meeting with friends, relatives, or neighbors?	1 (Daily) - 5 (Never)
	... helping out friends, relatives, or neighbors?	1 (Daily) - 5 (Never)
	... volunteer work in clubs or social services?	1 (Daily) - 5 (Never)
	... participation in political parties, municipal politics, or citizen initiatives?	1 (Daily) - 5 (Never)
	... going to church or attending religious events?	1 (Daily) - 5 (Never)

Table 6: The table presents the survey questions that were used to construct the integration target "social activity". In line with Aksoy et al. [8], our approach takes into account social interactions with both natives and non-natives, while excluding interactions with people from the same country of origin.

Target	ROC-AUC	PR-AUC	Brier-Score	F-Score	Precision	Recall
Employment	0.65	0.22	0.13	0.35	0.23	0.70
Social Activity	0.59	0.36	0.20	0.46	0.35	0.68

Table 7: The table reports the average performance of the integration prediction models across states for both target variables (see Section 4). The following probability thresholds were applied to measure the F-score, Precision, and Recall: 15% for employment, based on the mean employment rate, and 28% for social activity, based on the mean social activity rate. Lower performance results were expected since we replicated the tool, and deliberately avoided to fine-tune it. Our goal was not to optimize the tool, but to reconstruct and analyze it as-is transparently.

	Employment			Social Activity		
Group	Mean Outcome	Mean Prediction	Gain	Mean Outcome	Mean Prediction	Gain
Sex						
... Female	0.05	0.17	2.32	0.18	0.38	1.13
... Male	0.24	0.41	0.76	0.36	0.49	0.35
Country of origin						
... Syria	0.11	0.25	1.17	0.30	0.39	0.30
... Iraq	0.08	0.30	2.78	0.19	0.43	1.23
... Afghanistan	0.17	0.28	0.66	0.33	0.43	0.30
... Eritrea	0.33	0.38	0.15	0.21	0.55	1.60
Religious affiliation						
... Islamic religious denomination	0.08	0.47	4.81	0.32	0.47	0.48
... Shiite religious denomination	0.08	0.25	1.96	0.33	0.44	0.31
... Sunnite religious denomination	0.14	0.26	0.92	0.29	0.42	0.45
... Other religious denomination	0.07	0.34	3.56	0.22	0.41	0.80
... No religious denomination	0.24	0.31	0.28	0.20	0.39	0.94
Vocational training						
... No	0.16	0.29	0.83	0.27	0.43	0.62
... Yes	0.14	0.38	1.78	0.36	0.47	0.33
Education level (cat: high)						
... No	0.15	0.30	1.06	0.28	0.43	0.56
... Yes	0.15	0.34	1.21	0.37	0.48	0.28

Table 8: Gains from algorithmic assignment differ significantly across groups, particularly for those with lower mean outcomes. Gains are calculated as relative differences, defined as the difference between predicted and true outcomes, divided by the true outcome. Only groups with more than 5% observations in the test data are reported (see Table 5).