Beyond Proxy Variables: Extending Refugee Allocation Algorithms for Equitable Predictions

Clara Strasser Ceballos [1,*], Marcus Novotny [1,2,*], Christoph Kern[1,2]
[1] LMU Munich, Germany
[2] Munich Center for Machine Learning (MCML)
Clara.StrasserCeballos@stat.uni-muenchen.de, Marcus.Novotny@stat.uni-muenchen.de,
Christoph.Kern@stat.uni-muenchen.de

Abstract

Extensive social science research has long documented how individual- and location-level factors influence refugee integration. A more recent, smaller line of computational research has introduced algorithmic tools that predict refugees' integration outcomes across potential resettlement locations to inform placement decisions. However, these tools, currently being piloted in several countries, raise major concerns. They rely on a narrow set of predictors, many of which are considered protected attributes under anti-discrimination law, while omitting significant findings from explanatory migration research that could improve model predictions and reliability. Against this background, we draw on a systematic review of empirical migration research and comprehensive refugee panel data in Germany to improve the algorithmic modeling of refugees' integration outcomes. Specifically, we develop models that integrate and test a wide range of migration research variables to predict the economic integration of refugees arriving in Germany in 2016 and 2017. We compare our extended models to existing baselines from the algorithmic matching literature, evaluating both classification performance and fairness. Our results demonstrate that substituting proxy features with theory-driven variables which could be surveyed at arrival can considerably improve both accuracy and fairness without adversely impacting downstream allocation performance. We conclude that integrating insights from empirical migration research is essential for developing more reliable and robust algorithmic matching tools.

Appendix & Code — https://osf.io/t2s6y/

1 Introduction

Refugees who have been forced to flee their country of origin and are unable to return due to war, conflict, or fear of persecution, are often resettled from a country of asylum to a third host country that offers them international protection and eventually permanent residence (UNHCR 2024a,b).

Upon arrival in the host country, the successful integration of refugees into society becomes a key challenge (Ager and Strang 2008). Integration, a concept with contested definitions among scholars (Castles et al. 2003), can be understood as a two-way process in which both refugees and members of the host society adapt to each other (European Commission 2025). As such, it is not limited to a single dimension, but may include interrelated dimensions like economic, social, psychological, political, linguistic, and navigational (Harder et al. 2018).

To gain a better understanding of refugee integration, migration studies have developed theory-driven quantitative models that examine how both individual and initial placement location factors influence integration outcomes. A large number of these studies have focused on the economic integration of refugees, commonly measured by employment status (Hannafi and Marouani 2023). Research findings highlight that refugee integration cannot be fully explained by socio-demographic characteristics alone, such as sex (Kosyakova, Salikutluk, and Hartmann 2023) or education level (Brücker et al. 2024). Instead, it is shaped by a much broader range of individual and contextual factors, including the presence of children (Hans 2024), health status (Gambaro, Neidhöfer, and Spiess 2021), type of initial accommodation (Mendola, Parroco, and Li Donni 2024), or local conditions, such as unemployment rates (Aksoy, Poutvaara, and Schikora 2023) and availability of integration courses (Kanas and Kosyakova 2023).

A recent, separate line of research aims to improve refugee integration by informing more effective resettlement decisions. In many countries, designated resettlement agencies typically assign arriving refugees to specific locations (e.g., states or municipalities). However, these decisions are usually made either (quasi)randomly and/or based on administrative capacity constraints (BAMF 2024; RCUSA 2024; Swiss Refugee Council 2024), with limited consideration of how placement decisions affect refugees' integration (Åslund and Rooth 2007; Jones and Teytelboym 2018; Martén, Hainmueller, and Hangartner 2019). To address this, researchers have proposed algorithmic approaches that combine machine learning with matching techniques to assign refugees to presumed optimal resettlement locations upon arrival (Bansak et al. 2018a; Ahani et al. 2021a). This approach is motivated by the idea of refugee–location synergies, i.e., the notion that certain locations may be better suited to specific groups of refugees and that they can thus improve prospects of successful (economic) integration.

Two prominent refugee-location matching tools are GeoMatch (Bansak et al. 2018a; Immigration Policy Lab 2025)

[*] These authors contributed equally.

and Annie™ Moore (OPSI 2020; Ahani et al. 2021a), which have been piloted in the U.S., Switzerland, and the Netherlands (Global Compact on Refugees 2024; OPSI 2020). Both operate in two layers: (1) a *prediction* layer, where individual-level characteristics of refugees are used to predict their employment probabilities across resettlement locations upon arrival; and (2) a *matching* layer, where refugees are assigned to locations based on these predictions to maximize overall integration outcomes, subject to location-specific capacity constraints. The tools are introduced in more detail by Strasser Ceballos and Kern (2025) and in Appendix A.

Both tools draw on a narrow set of features for their model predictions, many of which are considered protected attributes under anti-discrimination law (Simson, Pfisterer, and Kern 2024). These variables, such as *sex*, *country of origin*, or *religious affiliation*, can be understood as *proxy variables* for contextual factors such as care obligations, family ties, or work histories. However, these proxies are at best error-prone and at worst discriminatory, as they cannot be affected by the individual (Jacobs and Wallach 2021). Stated differently, with the current reliance on demographic features, prediction models will differentiate refugees' integration outcomes predominantly based on their membership in different social groups, rather than learning the underlying mechanisms that lead to different integration likelihoods. Next to fairness implications, this also leads to questions of model robustness and generalizability, as the inferred relationships between proxy variables and integration outcomes may vary between different (sub)populations of refugees and over time. Furthermore, variables such as *country of origin* represent only one possible operationalization of broader, more complex concepts like *national origin* or *ethnicity*, posing complex measurement problems where each specific operationalization can lead to different fairness outcomes (Jaime and Kern 2024).

Ultimately, placement decisions in algorithmic matching tools depend on the underlying predictive models. However, to the best of our knowledge, no prior studies have: (1) systematically analyzed the importance of demographic predictors or justified their inclusion over other integration factors studied in migration research[1]; (2) extended the range of predictors beyond the few (nine to eleven) features included in current tools to improve model accuracy and fairness; and (3) evaluated group fairness of predictions, especially regarding commonly included protected attributes.

Contributions. Against this background, our study contributes in three key aspects. First, we conduct a systematic literature review of empirical migration studies on the economic integration of refugees, cataloging all relevant variables from prior literature. We then develop holistic predictive models that incorporate these insights. By training the models on a longitudinal survey of refugees in Germany, we study both their explanatory and predictive power (see Shmueli 2010). Second, we evaluate our extended predictive models against the feature set used in previous implementations of the GeoMatch tool, the most widely used algorithmic refugee-location matching tool to date (Global Compact on Refugees 2024), in terms of performance, group fairness, and variable importance. Third, we examine how substituting protected attributes with theory-driven behavioral variables identified in migration research affects model fairness.

Our empirical analysis yields several key findings. First, our extended models, which incorporate a rich set of predictors from quantitative migration research, consistently outperform the GeoMatch baseline predictor set across various performance metrics. Even the model specification limited to variables plausibly observable at or before arrival achieves an absolute improvement in ROC-AUC of 4%. Second, our fairness evaluations, using fairness notions tailored to the refugee-location matching context (Makhlouf, Zhioua, and Palamidessi 2021), show that the baseline GeoMatch model exhibits pronounced fairness disparities: women, Iraqi-born refugees, and individuals with less common religious affiliations experience substantially lower true-positive rates within the categories of these attributes. Third, our feature importance analysis highlights the predictive value of several additional variables. Some of these can be surveyed at the time of arrival, such as child-related information (e.g., presence of *children*, or *childcare in hours*) and employment-related factors (e.g., previous *full-time work experience*, stated *intention to work*). Others typically become observable shortly after arrival, such as the *type of accommodation* or *language proficiency*. In addition, contextual information about the resettlement location, such as the presence of *migrant networks*, also partially predicts integration outcomes. Finally, we demonstrate that replacing protected attributes with substantive, theory-driven variables reduces these fairness gaps, as the model then bases its predictions on more reliable behavioral measures, without compromising predictive performance or downstream allocation. This suggests that moving beyond proxy-based "shortcuts" can lead to more equitable and accurate matching tools.

Implications. Based on our systematic literature review and empirical findings, we highlight the value of integrating insights from social science research – in this case, quantitative migration research – into computational research and machine learning practice. In the context of algorithmic matching tools such as GeoMatch, we underline the potential to improve not only model performance, but also group fairness by incorporating theoretically-motivated variables. This is especially critical given that these algorithmic tools affect vulnerable populations and must be held to high standards of fairness and accountability.

Our paper is structured as follows. We first present our systematic literature review (Section 2). Further, we introduce our analytical approach (Section 3) and results (Section 4). We conclude with a detailed discussion (Section 5).

2 Systematic Literature Review

We conducted a systematic literature review (SLR) of quantitative empirical studies that examine how individual- and location-level characteristics are associated with refugees' economic integration outcomes. Two important consid-

[1]Annie™ Moore reports on the Gini importance of their basic feature set (Ahani et al. 2021b).

erations guided our approach: (1) refugee integration is highly context-specific (Phillimore 2021), and (2) meaningful quantitative analysis requires reliable data. Consequently, we focused our research on the German context for two primary reasons. First, the resettlement process of refugees in Germany follows a (quasi)random allocation mechanism. This makes Germany a strong case for empirical research, as it reduces (self-)selection bias in empirical analysis (Schilling and Stillman 2024). Second, Germany has the largest representative longitudinal survey of refugees in Europe: the IAB-BAMF-SOEP (Schilling and Stillman 2024). The survey has been conducted annually since 2016, and provides a rich source of individual-level data to assess integration trajectories over time (Brücker et al. 2025).

Methodology. We conducted a SLR following established guidelines, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page et al. 2021), and the recommendations by Carrera-Rivera et al. (2022). Our search process entailed four steps:

1. **Research objective definition:** Identify and analyze studies that have quantitatively examined the economic integration of refugees in Germany.
2. **Keyword search specification:** IAB-BAMF-SOEP *AND* integration.
3. **Source selection:** Scopus, Science Direct, Springer, Web of Science, and OpenAlex.[2]
4. **Inclusion criteria specification:**
   - Time frame: 01/2016[3] - 05/2025
   - Language: English and German
   - Publication type: Short reports and articles in journals, conferences and books.
   - Studies examining any form of economic integration.[4]
   - Studies employing empirical methods to analyze the relationship between factors and economic integration, using the IAB-BAMF-SOEP survey data for analysis.

We identified 690 articles in the initial search. After removing 131 duplicates, 559 articles were screened for eligibility by one reviewer and verified by a second. Ultimately, 25 studies met all inclusion criteria – quantitatively analyzing economic integration using the IAB-BAMF-SOEP – and were included in the final review. A detailed overview of the selection process is provided in Appendix B.

Results. Table 1 summarizes the relevant studies identified in our SLR. We extracted two key pieces of information from each study: (1) the variables included in the analysis; and (2) those found to be statistically significant for economic integration. For comparison, we also present the set of predictors used by the two existing algorithmic refugee-location matching tools, GeoMatch and Annie™ Moore.

To facilitate interpretation, we categorized all variables into three conceptual levels: *pre-arrival*, *post-arrival*, and *location*. These classifications are proposed by us rather than by the reviewed papers, and are meant to aid interpretation and practical application.

- *Pre-arrival* variables refer to refugees' individual-level characteristics that are either fixed (e.g., country of origin), were acquired before migration (e.g., education level), or could, in principle, be surveyed at the time of arrival (e.g., intention to work).
- *Post-arrival* variables include individual-level characteristics that evolve during the refugee's integration process (e.g., German language skills and asylum status), or circumstances that arise only after arrival (e.g., type of first accommodation and integration course participation).
- *Location* variables capture characteristics of the resettlement location (e.g., unemployment rate).

The reviewed studies differ in terms of their operationalization of economic integration (e.g., employment status or net income), sample populations (e.g., all, only female, or only Syrian refugees), and methodological approaches (e.g., propensity score matching or survival analysis). Despite these differences, a number of consistent findings emerge.

All predictors used in applications of the algorithmic tools are limited to pre-arrival characteristics. While many of them, such as sex, age, marital status, country of origin, and education, are consistently identified as significant in existing migration studies (Meyer 2025b; Brücker et al. 2024; Mendola, Parroco, and Li Donni 2024), additional relevant factors at this level are often overlooked. Notably, work experience, despite its well-documented and repeated association with economic integration (Bürmann and Tsolak 2025; Kosyakova and Kogan 2024; Brücker et al. 2024), is absent from these models. Similarly, health status and the presence of children are frequently identified as influential (Kosyakova, Salikutluk, and Hartmann 2023; Gambaro, Neidhöfer, and Spiess 2021; Kosyakova et al. 2021), yet are only incorporated in Annie™ Moore and not in GeoMatch.

Post-arrival variables are currently not considered in refugee matching tools. However, migration research consistently highlights their significance for economic integration outcomes. For instance, Meyer (2025a) finds that refugees living in shared accommodation are less likely to be economically integrated than those living in private housing. Kosyakova and Brenzel (2020) show that prolonged asylum procedures reduce refugees' employment prospects, emphasizing the impact of legal and procedural factors on economic integration. Additionally, a number of studies underline the importance of social networks. Kosyakova and Kogan (2024) highlight the importance of social contacts, particularly with co-ethnics, for employment outcomes, a finding supported by Meyer and Winkler (2023).

Finally, current algorithmic matching tools do not consider location characteristics as predictors.[5] Yet, a substantial body of migration research has consistently underlined

[2] To query the sources, we accessed their APIs with bibliometric tools (Rose and Kitchin 2019; Herrmann and Rose 2025).

[3] Survey publication year (Brücker, Rother, and Schupp 2017).

[4] In our SLR, we treat the terms economic, labor market and structural integration synonymously.

[5] As GeoMatch and Annie™ Moore estimate separate predictive models for each resettlement location (e.g., Swiss cantons), they implicitly treat location as a moderator for other variables, instead

Variable Categories	**Empirical Studies**																									**Tools**	
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**	**[12]**	**[13]**	**[14]**	**[15]**	**[16]**	**[17]**	**[18]**	**[19]**	**[20]**	**[21]**	**[22]**	**[23]**	**[24]**	**[25]**	**GM**	**AM**
Pre-Arrival																											
Character traits																				✓				✓*			
Children	✓	✓	✓*	✓	✓	✓			✓*	✓*	✓	✓*	✓*	✓	✓*	✓*			✓*	✓*	✓	✓*	✓*				✓
Education level	✓*	✓*	✓*	✓	✓	✓*	✓	✓*	✓*	✓*	✓	✓*	✓	✓	✓*	✓*	✓	✓	✓	✓*	✓	✓	✓*	✓*	✓*	✓	✓
Gender roles attitudes	✓									✓*					✓				✓*			✓					
Health status	✓		✓		✓						✓				✓*	✓			✓	✓*		✓*					✓
Intention to work						✓		✓*	✓														✓				
Language level (German)						✓*					✓				✓			✓		✓			✓*		✓*		
Language level (English)		✓*											✓*				✓*			✓*						✓	✓
Migration support friends/relatives						✓					✓	✓			✓								✓				
Religious affiliation											✓													✓		✓	
Socio-demographics	✓*	✓*	✓*	✓	✓	✓	✓	✓*	✓*	✓*	✓	✓*	✓*	✓	✓*	✓*	✓*	✓	✓*	✓*	✓	✓*	✓*	✓*	✓*	✓	✓
Traumatic experiences												✓			✓												
Work experience	✓	✓*	✓		✓	✓*			✓*			✓	✓*		✓*			✓	✓*		✓	✓	✓*		✓*		
Other								✓				✓			✓								✓	✓			
Post-Arrival																											
Accommodation type	✓		✓*	✓			✓	✓*	✓				✓		✓	✓				✓			✓	✓			
Asylum status	✓	✓*	✓	✓	✓			✓*	✓	✓		✓*		✓	✓	✓	✓*		✓		✓*	✓*	✓*		✓*		
Credential recognition					✓				✓*						✓							✓					
Integration course participation		✓*	✓*		✓				✓	✓	✓	✓				✓			✓			✓		✓	✓*		
Labor market course participation	✓		✓*		✓*				✓							✓						✓*			✓*		
Language course participation			✓*		✓				✓*	✓					✓	✓			✓*			✓*	✓*		✓*		
Language level (German)	✓*	✓								✓*			✓*		✓	✓*	✓*		✓*			✓*		✓*			
Social contact co-ethics	✓												✓*		✓*	✓*						✓*					
Social contact natives	✓	✓*								✓		✓			✓*	✓*			✓			✓*					
Social contact other		✓											✓*		✓*							✓					
Support job seeking						✓*													✓*			✓*					
Xenophobia perception	✓							✓*	✓						✓												
Other		✓						✓	✓			✓	✓		✓		✓		✓	✓		✓	✓				
Location																											
Attitudes towards migrants											✓*	✓															
GDP per capita							✓							✓		✓*	✓										
Population density							✓					✓			✓		✓										
Share of foreigners				✓			✓				✓	✓		✓	✓	✓	✓	✓*		✓			✓				
Unemployment rate				✓	✓		✓*				✓*	✓*			✓	✓*	✓			✓			✓				
Voting behavior				✓			✓				✓	✓		✓						✓							
Other							✓					✓		✓	✓	✓	✓			✓							

Table 1: **Unlike GeoMatch and Annie™ Moore, empirical migration research incorporates a broad set of predictors at the pre-arrival, post-arrival, and location levels when analyzing economic integration.** The table provides an overview of 25 empirical studies on the economic integration of refugees in Germany, based on the IAB-BAMF-SOEP survey. The studies, identified through a systematic literature review, are arranged chronologically by year of publication (left to right). The predictor sets of matching tools GeoMatch (GM) (Bansak et al. 2018b) and Annie™ Moore (AM) (Ahani et al. 2021b) are shown in the two rightmost columns. Socio-demographic characteristics include: age at arrival, sex, family status, and country of origin. A check mark (✓) indicates the variable was included in the study; an asterisk (*) indicates that the variable was statistically significant at any level ($p < 0.1$) in the main model specification of the study. Control variables that are included across studies but are not listed in the table are: immigration year and assigned federal state and/or municipality. References correspond to the following enumerated columns: Meyer (2025b) [1], Bürmann and Tsolak (2025) [2], Meyer (2025a) [3], Khalil and Tjaden (2025) [4], Bredtmann and Höckel (2024) [5], Kosyakova and Kogan (2024) [6], Schilling and Stillman (2024) [7], Mendola, Parroco, and Li Donni (2024) [8], Brücker (2024) [9], Hans (2024) [10], Aksoy, Poutvaara, and Schikora (2023) [11], Kanas and Kosyakova (2023) [12], Hannafi and Marouani (2023) [13], Tjaden and Spörlein (2023) [14], Kosyakova, Salikutluk, and Hartmann (2023) [15], Meyer and Winkler (2023) [16], Tsolak and Bürmann (2023) [17], Battisti, Peri, and Romiti (2022) [18], Salikutluk and Menke (2021) [19], Gambaro, Neidhöfer, and Spiess (2021) [20], Strazzeri (2021) [21], Kosyakova et al. (2021) [22], Kosyakova and Brenzel (2020) [23], Hahn et al. (2019) [24], Kosyakova and Sirries (2017) [25].

of explicitly including regional covariates. This has no theoretical grounding in prior literature and is also data-inefficient, as many predictor–outcome relationships likely generalize across locations.

the role of resettlement location factors – such as local unemployment rates (Aksoy, Poutvaara, and Schikora 2023), the share of foreigners (Battisti, Peri, and Romiti 2022), public attitudes towards immigrants (Schilling and Stillman 2024), and supply of language courses (Kanas and Kosyakova 2023) – in shaping refugees' integration.

Discussion. Our SLR reveals a key limitation of existing algorithmic matching tools: their reliance on a narrow set of predictors, restricted entirely to pre-arrival variables. Even within this category, several empirically important factors, such as prior work experience, are notably absent, despite being consistently identified as relevant in migration studies. While excluding post-arrival variables is methodologically consistent with the tools' goal of generating predictions at arrival, it remains crucial to test their predictive value. Some of these factors may be anticipated through early indicators that could feasibly be collected upon arrival. A further shortcoming of current tools is the lack of location-level variables. This is surprising, as most are publicly available, easy to integrate, and have repeatedly been shown to significantly impact integration.

To address these limitations and encourage closer collaboration between social and computer science, our study incorporates empirical migration research insights into algorithmic matching. We demonstrate that this approach can improve both model accuracy and reliability.

3 Empirical Analysis: Evaluating Refugee Integration Predictions

We empirically investigate the prediction layer of algorithmic matching tools. To this end, we use German refugee data and simulate the application of the GeoMatch tool with different sets of predictors. In our case, a hypothetical state agency is assumed to generate individual employment predictions for refugees arriving in 2016 and 2017, based on historical patterns learned from 2014 to 2015 data.

Data Source

We draw on the latest version (v39) of the IAB-BAMF-SOEP Survey of Refugees in Germany (Brücker, Rother, and Schupp 2017; DIW 2024). This longitudinal survey, conducted annually since 2016 and integrated into the German Socio-Economic Panel (SOEP), provides rich information on refugees who arrived in Germany since 2013. To ensure representativeness, the survey draws random samples from the Central Register of Foreigners (AZR). As a result, it offers high-quality data for quantitative research (Aksoy, Poutvaara, and Schikora 2023) and has inspired similar data collection efforts in other countries (Ortlieb et al. 2024).

Target and Predictors

Target. We aim to predict economic integration, operationalizing it as a binary variable that indicates whether or not a refugee engages in any form of paid employment within three years of arrival in Germany.[6]

This three-year horizon is justified by two considerations. First, employment rates among refugees are typically low shortly after arrival, as integration into the labor market can take several years (Brücker 2024). Second, according to German policy, refugees are restricted to remain in their initial federal state of assignment for at least three years (Baba et al. 2024). Our approach follows prior work, such as the Swiss GeoMatch application, which also adopts a three-year outcome window (Bansak et al. 2018a).

Predictors. We reconstruct the variables identified in prior migration research (see Table 1) as well as those used by the GeoMatch tool, following our three-level categorization into *pre-arrival*, *post-arrival*, and *location-level* variables (see Section 2).[7] In total, we construct a comprehensive set of predictors, comprising 24 *pre-arrival*, 19 *post-arrival*, and 14 *location-level* variables. These also include additional behavioral variables that have not been considered in prior research, but could capture refugee characteristics better than existing protected attributes (e.g., religious event attendance frequency vs. religious affiliation).

At the *location-level*, we adopt the set of variables recommended by migration studies at the federal state level. These include GDP per capita (Statistik Portal 2025), unemployment rate (DESTATIS 2025a), population density (DESTATIS 2025c), voting behavior (measured by the share of second votes for major parties in the latest federal election) (INKAR 2025), share of foreigners (INKAR 2025), and public attitudes toward migrants (measured via survey-based sentiment indicators) (ESS 2025). We also extend this list by including the size of local foreign-national communities (DESTATIS 2025b), the total number of reported crimes (PKS 2025), and the number of integration courses offered by the Federal Office for Migration and Refugees (BAMF-NAvI 2025). We incorporate location-level predictors based on each refugee's allocated federal state, and lag them by one year relative to the immigration year.

Analytical Setup

Train-Test Split. We draw on the M3-M5 samples of the IAB-BAMF-SOEP refugee survey. The samples comprise 9,459 refugees (26,495 person-year observations) in Germany who have completed at least one annual survey between 2016 and 2022.

Given our objective – to assess how individual- and location-characteristics at arrival predict employment within three years – we restrict the sample to refugees who: a) are of working age, b) completed their first survey within one or two years of arrival, c) remained in the panel for at least three years after immigration, and d) were not employed at the time of their first survey.[8]

The final dataset consists of 5,477 refugees, who immigrated between 2014 and 2017. We split the data by immi-

[6]This includes full-time, part-time, marginal, or irregular employment, apprenticeships, vocational retraining, and internships.

[7]As no data were available, we excluded *credential recognition* and *integration course participation*.

[8]We exclude refugees already employed at the time of their first survey to avoid reverse causality. This is necessary as some post-arrival characteristics (e.g., social contacts) may both influence employment or result from being employed.

gration year, rather than randomly, to mimic the real-world allocation process. In a deployment scenario, newly arriving refugees are assigned to locations based on data learned from past refugee arrivals (Bansak et al. 2018a). In our case study, refugees in either 2014 or 2015 form the train set (n=4,520), while those arriving in 2016 or 2017 form the test set (n=927). Although all refugees were first surveyed in 2016 or 2017, we assume that responses from refugees in the train set still approximate conditions at their time of arrival.

Summary statistics for the train and test sets are provided in Appendix C, Table 1. Distributions are comparable between both sets, with 17-18% of refugees finding employment within three years. The sample is approximately evenly divided between male and female refugees. The largest countries and regions of origin are Syria (roughly 50%), followed by Iraq and Afghanistan. All other countries appear in much smaller numbers, and are aggregated at a regional level into Africa, Southwest Asia, Europe and Others.[9]

Table 2 in Appendix C further breaks down employment rates by protected attribute and train-test set. Men are substantially more likely to find work (25-29%) than women (5-8%). Employment rates also vary by country / region of origin, ranging from 12-15% for Iraqis compared to 22-28% for African-originating refugees. Religious affiliation shows smaller employment differences and no clear patterns across train-test split. This uneven ground-truth distribution underscores the importance of evaluating both overall predictive accuracy and group fairness in our models.

Prediction Models. We implement a standardized classification machine learning workflow with stochastic gradient boosting machines (Friedman 2001). Our setup closely follows the GeoMatch approach (Bansak et al. 2018a), but uses single-pooled instead of state-specific models. This allows us to include location-level predictors and efficiently leverage the full dataset. We fit eight model specifications, organized into two categories (see Table 2). Models A1-A4 incrementally add predictor groups from Table 1, ultimately including the full set of variables identified in the SLR. Models B1-B4 follow the same progression but exclude three salient protected attributes: sex, country / region of origin, and religious affiliation.[10]

Additionally, we define two null models as naive baselines. The first *Mode* model assigns the most frequent outcome from the train data to all cases (i.e., no employment). The second *Weighted* model assigns employment outcomes randomly to individuals according to the employment rate in the train set (i.e., 17.3%).

Hyperparameter optimization is conducted via Bayesian (Model-Based) Optimization (Garnett 2023), using Precision-Recall AUC (PR-AUC) as the objective metric. We employ nested resampling, with an initial single hold-out split into train and test sets. Within the train set, we perform three-fold cross-validation for hyperparameter tuning, using stratified resampling by target variable and sex. Hyperparameter search spaces and tuning results are documented in Appendix C, Table 3. The respective best models from hyperparameter tuning are then retrained on the full train set and used to generate predictions for the test set. To address class imbalance and ensure comparability, we employ a "top K" thresholding strategy, choosing cutoffs that yield exactly 160 positive predictions. This matches the expected number of employed cases in the test set (n=927) when assuming train set employment rates (17.3%).

Category	Models		Predictors
A) Full Models	A1)	Baseline	10
	A2)	Pre-Arrival (incl. A1)	24
	A3)	Post-Arrival (incl. A2)	43
	A4)	Location (incl. A3)	57
B) No Protected Variables	B1)	Baseline	7
	B2)	Pre-Arrival (incl. B1)	21
	B3)	Post-Arrival (incl. B2)	40
	B4)	Location (incl. B3)	54
Null Models	Mode		/
	Weighted		/

Table 2: **Model specifications.**

Evaluation

Classification Performance. We evaluate our models using standard measures, prioritizing the threshold-independent metrics ROC-AUC and PR-AUC. As PR-AUC is sensitive to class imbalance, it ensures that performance gains are not driven by the majority "unemployed" class. We also include threshold-dependent metrics, Accuracy and Balanced Accuracy, for comparison with prior algorithmic matching benchmarks and "real-world" performance evaluation, which would also require a threshold. Lastly, we report True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), and False Negative Rate (FNR).

Fairness. We evaluate fairness for three protected attributes – sex, country / region of origin, and religious affiliation – because they a) are legally protected under EU anti-discrimination law, b) have been persistently linked to labor market discrimination in Europe (Zschirnt and Ruedin 2016; Bishu and Alkadry 2017), and c) show sufficient representation in our data to yield reliable subgroup estimates.

Following the selection scheme by Makhlouf, Zhioua, and Palamidessi (2021), we adopt two group fairness metrics: Equalized Odds Difference and Equal Opportunity Difference. Equalized Odds demands that both TPR and FPR are equal across groups; its distance measure (EOD) is the average of the maximum TPR and FPR difference across groups. Equal Opportunity relaxes this to require equality only in TPRs; its distance measure (Δ TPR) is the maximum TPR difference across groups.

We focus on these metrics since correctly identifying refugees who will find employment (i.e., minimizing false

[9]We follow an aggregated version of the region coding system specified by the United Nations (1999).

[10]As prior models like GeoMatch were built around individual-level variables, we add location variables last, despite them being available before arrival. Due to their limited predictive power, changing this order of inclusion did not meaningfully affect results.

negatives) is crucial in the algorithmic matching context: a missed positive prediction denies an individual access to an optimal resettlement match. To complement these threshold-based metrics, we examine "balance for the positive class" by comparing the full distributions of predicted employment probabilities across protected groups in an auxiliary analysis (Makhlouf, Zhioua, and Palamidessi 2021). While differing base rates rule out identical distributions, a fair model should ensure that qualified individuals regardless of subgroup receive similarly high scores.

Feature Importance. To uncover how extending our set of predictors and substituting protected attributes affect model behavior, we assess feature importance via the relative-influence metric, quantifying each feature's contribution to in-sample loss reduction (Friedman 2001). Importance scores are computed on the train set and normalized to sum to one, facilitating direct comparison across models.

Downstream Allocation Impact. We evaluate the impact of our extended predictor sets on allocation outcomes. Following the GeoMatch framework, we simulate optimal alternative placements for refugees in our test set (Bansak et al. 2018a). To this end, we generate 16 state-specific predicted employment probabilities per refugee and solve a capacitated assignment problem that assigns each refugee to the state maximizing overall average employment probability, while preserving state-level quotas observed in the data.[11]

We compare the resulting *optimal predicted* average employment rate under this optimal allocation against two benchmarks: 1) the *actual* employment rate realized in the test set (18.1%), and 2) the *predicted*, model-specific employment rate under historical assignments, using each refugee's predicted probability for their actually assigned state. Benchmark 1) is adapted from the original paper and requires well-calibrated probabilities that reflect real-world likelihoods, while 2) controls for potential distribution shift and calibration error. This is relevant as our model was optimized only for discrimination between cases (i.e., PR-AUC), and gradient-boosted machines often yield poorly calibrated scores that are good for ranking but biased in absolute terms (Niculescu-Mizil and Caruana 2005). We thus also test several calibration approaches, including isotonic, beta, platt, and linear rescaling for this analysis.

4 Results

Classification Performance

Table 3 compares performance across our eight model specifications and two null models. The GeoMatch baseline (A1) achieves a ROC-AUC of 0.73 and PR-AUC of 0.32, showing only modest predictive ability. The successive inclusion of pre-arrival and post-arrival features incrementally raises ROC-AUC to 0.77 and PR-AUC to 0.42 (A3), with corresponding gains in recall. However, we observe no further improvement for the Location model (A4). Accuracy and balanced accuracy improve only marginally and lag behind the *Mode* null model, a consequence of the severe class imbalance in employment and the difficulty of correctly predicting outcomes – with TPR < 0.5 across models, each positive prediction necessarily worsens accuracy. Although always predicting the mode "no employment" label yields superficially high accuracy, it has zero utility in practice. In contrast, all model variants improve accuracy against the *Weighted* null model.

When removing protected attributes (Category B), we find that only the baseline specification (B1) exhibits a meaningful drop in classification performance (from 0.73 to 0.68 ROC-AUC / 0.32 to 0.26 PR-AUC). Crucially, supplementing removed attributes with pre-arrival variables (B2) restores performance close to the full model level, with similar effects for Post-Arrival (B3) and Location (B4) models.

Fairness

Table 4 shows fairness evaluation results across protected attributes. In full models A1–A4, fairness is uniformly poor. Min(TPR) for women is essentially zero (0.00–0.02), with Δ TPR $\approx$ 0.45–0.58, meaning that (almost) no woman, but every second man is correctly classified as finding employment. Adding features does not alleviate this gap. Fairness for country / region of origin also remains inadequate: Min(TPR) for the worst-off subgroup (Iraq) is 0.04–0.22, with Δ TPR $\approx$ 0.37–0.62 to African refugees, and only slightly lower gaps compared to Syrian and Afghan refugees. Religion also shows very low Min(TPR)=0.13 for the worst-off group ("Other") in the GeoMatch-Baseline (A1). However, fairness scores improve in models A3-A4 for both religion and country / region of origin.

Evaluating our model specifications without protected variables reduces disparities for sex and country / region of origin, especially in extended model configurations (B2-B4). Even the baseline model (B1) already achieves improved metrics, although at the cost of decreased classification performance. Adding pre- (B2) and post-arrival features (B3) further improves Min(TPR), Equal Opportunity, and Equalized Odds across protected attributes, while our Location model (B4) achieves no gains. Overall, the largest fairness improvements occur from the removal of protected attributes, followed by incorporating substantive pre- and post-arrival predictors. While fairness metrics depend on threshold, these improvements persist for all models and across the full range of sensible thresholds, indicating that they are not unduly influenced by our "top K" strategy.

We also inspect employment probability distributions across models and sex in Appendix C, Figure 2. We find that female employment probabilities are strongly below male probabilities in models A1-A4, with interquartile ranges not even overlapping. In comparison, models excluding protected variables show a lot wider probability distributions for women, better reflecting their true employment rates.

The null models underscore the trade-offs in classification performance and fairness. The *Mode* classifier is perfectly fair, while the *Weighted* null model is closer to the notions of Equal Opportunity or Equalized Odds than all our other model specifications. However, both are entirely uninformative in practice.

[11] We adapt the original GeoMatch code base, and refer interested readers to Bansak et al. (2018a) for methodological details.

Category	Model	ROC-AUC	PR-AUC	Accuracy	Bal. Acc.	TPR	FPR	TNR	FNR
A) Full Models	A1) Baseline	0.73	0.32	0.77	0.60	0.34	0.14	0.86	0.66
	A2) Pre-Arrival	0.76	0.35	0.78	0.63	0.38	0.13	0.87	0.62
	A3) Post-Arrival	0.77	0.42	0.81	0.66	0.44	0.11	0.89	0.56
	A4) Location	0.77	0.41	0.80	0.66	0.43	0.12	0.88	0.57
B) No Protected Var.	B1) Baseline	0.68	0.27	0.76	0.58	0.31	0.14	0.86	0.69
	B2) Pre-Arrival	0.74	0.34	0.79	0.63	0.39	0.12	0.88	0.61
	B3) Post-Arrival	0.76	0.41	0.80	0.65	0.42	0.12	0.88	0.58
	B4) Location	0.77	0.41	0.80	0.65	0.42	0.12	0.88	0.58
Null Models	Mode	0.50	0.18	0.82	0.50	0.00	0.00	1.00	1.00
	Weighted	0.50	0.18	0.72	0.50	0.17	0.16	0.84	0.83

Table 3: **Extending the predictor set improves model performance across various metrics.** The table reports classification performance for all model specifications.

		Sex			Country / Region			Religion		
Category	Model	Min(TPR)	ΔTPR	EOD	Min(TPR)	ΔTPR	EOD	Min(TPR)	ΔTPR	EOD
A) Full Models	A1) Baseline	0.00	0.45	0.39	0.04	0.62	0.48	0.13	0.40	0.28
	A2) + Pre-Arrival	0.02	0.47	0.39	0.09	0.58	0.46	0.13	0.45	0.27
	A3) + Post-Arrival	0.00	0.58	0.43	0.22	0.37	0.28	0.27	0.26	0.21
	A4) + Location	0.02	0.53	0.40	0.17	0.45	0.32	0.27	0.38	0.25
B) No Protected Var.	B1) Baseline	0.15	0.22	0.16	0.09	0.41	0.31	0.13	0.40	0.29
	B2) + Pre-Arrival	0.20	0.25	0.23	0.26	0.32	0.22	0.27	0.44	0.27
	B3) + Post-Arrival	0.24	0.24	0.20	0.43	0.17	0.14	0.27	0.32	0.26
	B4) + Location	0.22	0.27	0.22	0.35	0.26	0.19	0.27	0.38	0.29
Null Models	Mode	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Weighted	0.10	0.09	0.05	0.22	0.09	0.06	0.20	0.03	0.04

Table 4: **Fairness is uniformly poor across full models (A1-A4), but substantially improves when protected variables are removed (B1-B4).** The table presents fairness metrics for all model specifications. Min(TPR) indicates the TPR of the most disadvantaged group (*Female* for sex, *Iraq* for Country/Region, *Other* for Religion). Δ TPR is the maximum TPR difference between groups (*Equal Opportunity Difference*). EOD is the mean of the maximum differences in TPR and FPR (*Equalized Odds Difference*). TPR and FPR for all groups are presented in Appendix C, Tables 4 and 5.

Feature Importance

Full Model Specifications (A1-A4). Figure 1 displays relative feature importance across model specifications A1-A4. Several patterns emerge.

First, a small set of features captures most of the predictive signal and is already included in the GeoMatch baseline (A1). In every specification, the two leading predictors *sex* and the assigned *federal state* together account for one to two-thirds of total importance. *Age* also contributes strongly across models (∼6–16%). Several of the following behavioral variables are then only included in extended model specifications: e.g., refugees' *German proficiency* after arrival, *intention to work*, *childcare in hours*, and *full-time work experience*. Relative importance then quickly drops to 2% and below for variables outside the top 15.

Regarding protected variables, *sex* consistently ranks highest, contributing nearly half of in-sample loss reduction in A1 and still above 20% in A3–A4. *Country / region of origin* accounts for 2–5% across models and is also important, while *religious affiliation* never enters the top 20. However, once post-arrival variables are added, *religious events frequency* attains roughly 3%. Overall, we note that expanding the set of predictors already shifts the model's learned decision logic away from protected and demographic markers, such as *sex*, towards behavioral measures such as *childcare hours* and *work experience*.

Finally, location factors at arrival appear to play a minor predictive role. The only such variable to emerge, *share of migrant network* in the assigned federal state, ranks only 17$^{\text{th}}$. Moreover, the near-identical importance scores between A3 and A4 across shared features suggest that adding location-level data contributes little additional value beyond detailed individual-level predictors.

Without Protected Variables (B1-B4). Having already noted an importance shift from protected and demographic markers to behavioral measures when extending our set of predictors, we also compare full and reduced model specifications. For simplicity, Figure 2 contrasts feature importance in the full "Location" model with protected attributes (A4) against its counterpart without (B4). By definition, variable importance for *sex* and *country / region of origin* drops to zero in B4, since they are excluded from the model. We observe that their predictive influence is effectively redistributed onto more substantive behavioral measures. In par-

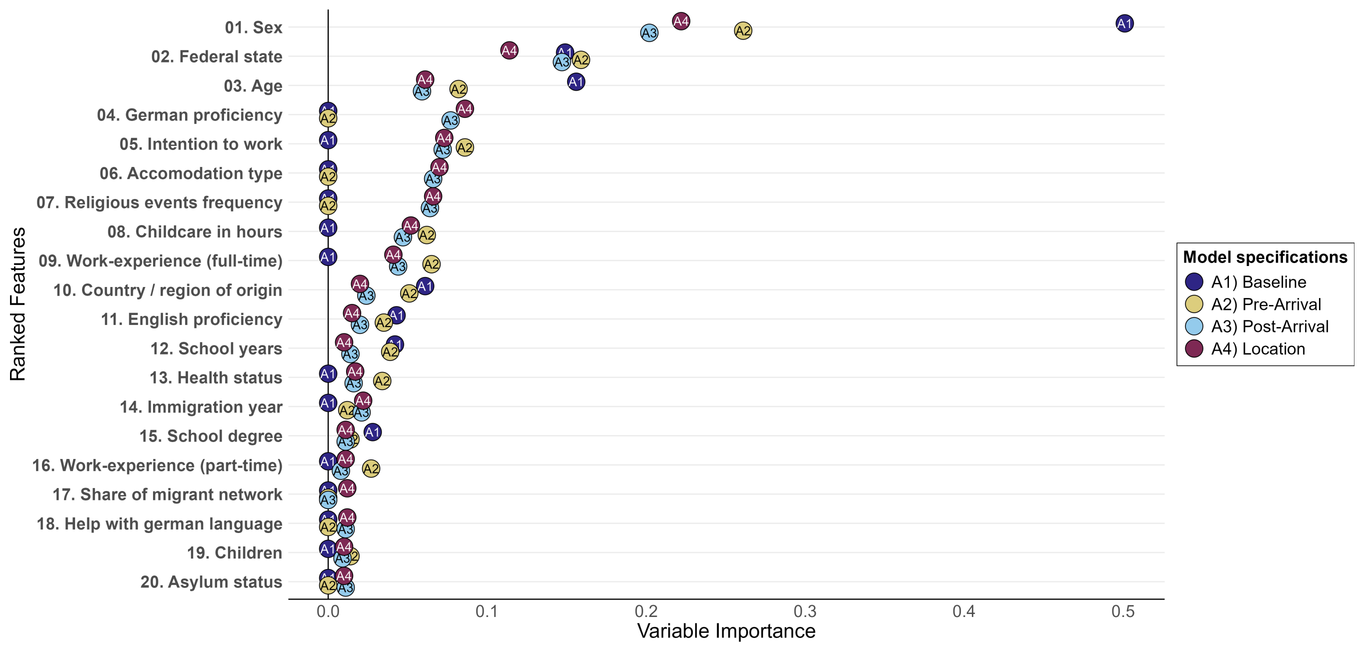

Figure 1: **While the top 3 features are present in all models, several of the top 10 features are missing from the baseline GeoMatch specification** (e.g., *intention to work* or *work experience*). The figure shows the top 20 features across model specifications A1-A4, ranked by mean relative importance (i.e., contribution to in-sample loss reduction). Features missing from model specifications are imputed with zero for visualization (e.g., *intention to work* for A1).

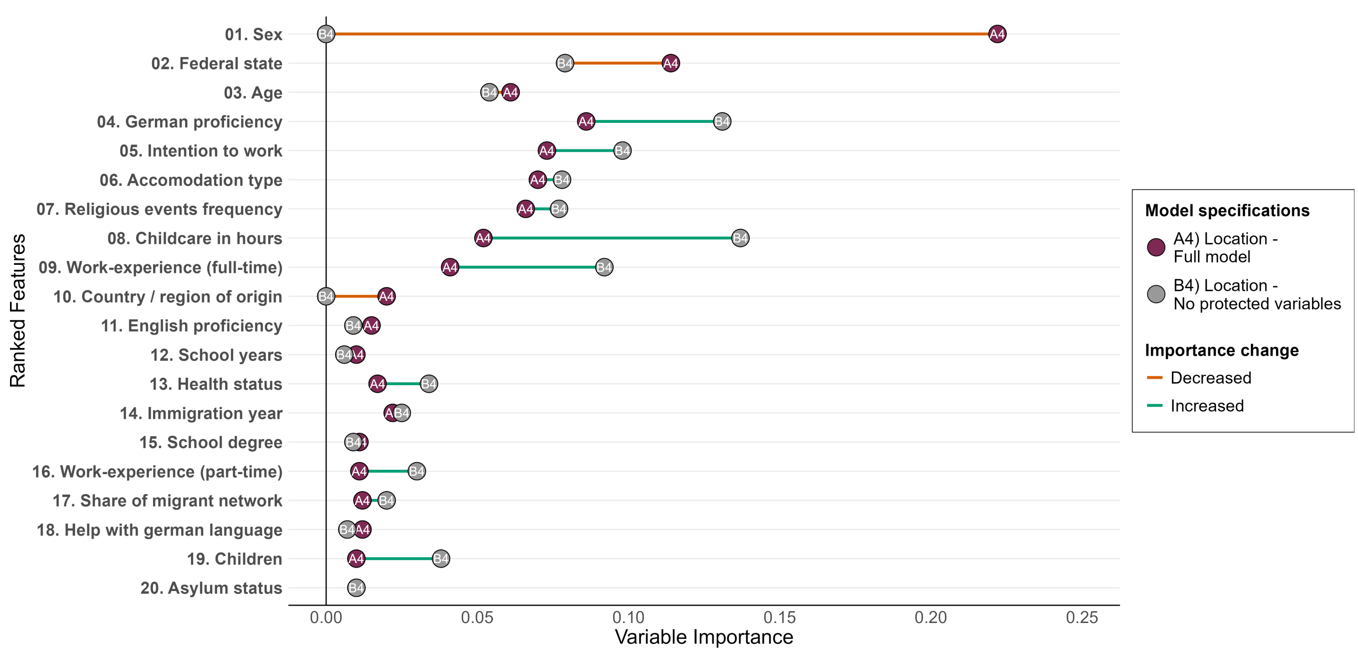

Figure 2: **When protected variables *sex*, *country / region of origin* and *religious affiliation* (not shown) are removed from the model, behavioral measures such as *german proficiency*, *childcare in hours* and *work-experience* rise in importance.** The figure compares feature importance between Location model specifications with (A4) and without (B4) protected variables.

ticular, *childcare in hours*, *full-time work experience*, and *German proficiency* after arrival gain most in B4, suggesting they may be the underlying drivers of the effects of protected attributes such as sex and country of origin in A4.

Downstream Allocation Impact

Finally, we compare predicted employment gains from the GeoMatch allocation algorithm for all models against two benchmarks: actual test-set employment rates (18.1%) and predicted model-specific employment rates given historical allocations. Testing several calibration approaches, we linearly rescale predictions to match the 17.3% train-set employment rate for comparison. The results are summarized in Appendix C, Table 6. Overall, we observe relative predicted employment gains between 8% and 22%. We find no clear pattern for the number of included variables or the exclusion of protected attributes. Hence, replacing protected attributes does not negatively impact downstream allocations.

5 Discussion

Resettlement agencies in several countries use algorithmic matching tools, such as GeoMatch, to predict refugees' integration likelihoods upon arrival and then assign them to locations (e.g., federal states) that maximize overall predicted outcomes. Although matching decisions are primarily based on predictions, there is surprisingly little research on the design, performance, and fairness of the underlying prediction models themselves. Against this background, we examine the prediction layer of the GeoMatch tool using German refugee data and highlight the importance of including meaningful behavioral predictors from migration research for both performance and reliability.

In the first part of our study, we conducted a systematic literature review of quantitative migration studies examining the role of individual- and location-level characteristics for refugees' economic integration. The review highlights the broad range of relevant variables identified in migration research. These extend far beyond the narrow set of individual-level, pre-arrival predictors –including a number of protected attributes and social categorizations – used by current algorithmic tools. Notably, several of the individual-level variables identified in the literature could feasibly be surveyed at arrival (e.g., prior work experience). Similarly, location-level variables could be sourced from administrative data (e.g., the size of local migrant networks).

In the second part of our study, we extended GeoMatch's predictor set by the variables identified in our systematic review. Incorporating these theoretically-motivated predictors yields incremental, yet consistent gains in overall accuracy: even the addition of pre-arrival variables alone meaningfully outperforms the GeoMatch baseline. Crucially, substituting protected attributes with this richer feature set also improves group fairness: While the standard GeoMatch predictor set results in disparate errors across protected groups, particularly for *sex* and *country / region of origin*, adding behavioral measures such as *work experience*, *childcare hours*, and *language proficiency* improves both accuracy and fairness. Feature-importance analyses confirm that our new predictors displace reliance on coarse demographic markers, focusing the model on predictors with clearer causal links to employment. Finally, we show that these adjustments come at no cost to downstream allocation: simulated matching under extended model specifications produces similar uplifts in predicted employment gains as observed for GeoMatch.

Naturally, our study comes with several limitations. Some of them are linked to our data and analytical design. First, we draw on a single-country sample with one train-test split, which may limit the generalizability of our findings to different country contexts or periods. We thus encourage future research to validate our findings in other settings. While our study of the German context may share similarities with other EU countries, such as the Netherlands, due to a similar migration and asylum policy (European Commission 2024), we particularly highlight the need for more research in the U.S. and Switzerland, where algorithmic matching tools are also currently being deployed. Second, we rely on survey data collected one to two years after a refugee's arrival, rather than on administrative records, which may introduce survey-related biases. When integrating other data sources, we also urge the consideration of refugees' personal resettlement preferences in evaluation, which have not yet been systematically surveyed but could positively impact integration outcomes. Third, while we used employment status to operationalize refugee integration, several other design choices are possible. We acknowledge that employment alone cannot capture the broad, multidimensional nature of refugee integration. Future research could therefore explore additional integration dimensions, such as social integration, or establish multidimensional indicators (see, e.g., Harder et al. 2018). Fourth, we are fully aware that removing protected attributes is not in itself a bias mitigation strategy, and that fairness is a multidimensional concept with many competing metrics. This may also include notions of individual fairness and multi-group (intersectional) fairness that critically extend beyond group fairness. Finally, while we are confident that most of the identified important variables are practically feasible to collect, we acknowledge that certain theoretically grounded predictors (e.g., personal attitudes or intentions) are difficult or ethically sensitive to measure at arrival, which may limit their real-world applicability. Collecting this data can also introduce measurement errors, e.g., when respondents give strategic answers or lack trust (Strasser Ceballos and Kern 2025). Hence, we advise practitioners to define minimally feasible yet comprehensive feature sets and choose data collection methods that maximize accuracy.

Overall, our findings highlight that improving refugee-location matching tools requires not only advances in data sciences but also sustained, interdisciplinary collaboration with the social sciences. We believe that teams comprising migration scholars and fairness-oriented machine learning researchers are best positioned to develop both accurate and equitable tools for refugee integration. Migration researchers can guide variable selection and account for both societal and labor market dynamics, while machine learning experts can develop robust evaluation strategies to ensure that the decision outcomes are monitored not only for performance, but also for fairness standards. Lastly, we hope that our work may also inspire future migration research. As statistically significant variables from explanatory research are not by default important predictors (Shmueli 2010), researchers could also explore further variables for their predictive power in this or related application scenarios.

More broadly, we highlight the importance of integrating insights across computer and social sciences to ensure that the algorithmic systems governing our lives are not only high-performing, but also socially reliable.

Ethical Statement

Our research raises three ethical considerations:

1. **Fairness notion choice:** To assess group fairness in our predictions, we selected several recognized fairness metrics based on established guidelines. However, we acknowledge that these metrics cannot capture the full range of fairness considerations, particularly in the context of refugee-location matching, which is a two-layered process. This complexity highlights the need for a more nuanced, context-sensitive approach to fairness in such settings.
2. **Multidimensional integration:** To study the predictive layer of the algorithmic matching tools, we limited our study to the economic dimension of integration, proxied by employment status. However, we recognize that integration is inherently multidimensional and that its various dimensions are deeply interconnected.
3. **Data privacy and confidentiality:** We relied on individual-level refugee data from the IAB-BAMF-SOEP survey, which requires a strict application process for data access. To safeguard privacy and confidentiality, we only report data aggregated at the group level.

Adverse Impact Statement

Our research may have three crucial unintended consequences:

1. **Misinterpretation of fairness through exclusion:** While we show that replacing protected attributes with theory-informed behavioral variables can reduce disparities in prediction outcomes, this should not be interpreted as a general recommendation to exclude protected attributes as a fairness mitigation strategy ("fairness through unawareness"). On the contrary, we emphasize the importance of having access to these attributes for fairness evaluations, as also predictors highly correlated with protected attributes may introduce subgroup disparities.
2. **Overlooking privacy and feasibility concerns:** While our findings highlight the value of a broad set of individual-level predictors, we acknowledge that collecting such data at arrival may not always be legally permissible or practically feasible. Implementation must account for privacy regulations, ethical standards, and administrative constraints.
3. **Individual agency:** Treating refugees as subjects to be matched to a location may undermine their agency and dignity by reducing them to data points and denying them a voice in location choices. Reliance on (black-box) machine-learning models may also erode trust and accountability in these decision-processes, especially when they are insulated from human oversight and lack clear avenues for appeal. Given that the status quo of (quasi)random allocation is not much more equitable in this regard, we encourage research that explicitly studies refugees' preferences and their impact on integration outcomes.

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 was supported by the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

We would like to thank our two research assistants, Oleksandra Bogdanova and Fedor Miasnikov, for their valuable help in data collection, and the anonymous reviewers for their insightful comments.

References

Ager, A.; and Strang, A. 2008. Understanding Integration: A Conceptual Framework. *Journal of Refugee Studies*, 21(2): 166–191.

Ahani, N.; Andersson, T.; Martinello, A.; Teytelboym, A.; and Trapp, A. C. 2021a. Placement Optimization in Refugee Resettlement. *Operations Research*, 69(5): 1468–1486.

Ahani, N.; Andersson, T.; Martinello, A.; Teytelboym, A.; and Trapp, A. C. 2021b. Supplementary Material: Placement Optimization in Refugee Resettlement. *Operations Research*, 69(5): 1468–1486.

Aksoy, C. G.; Poutvaara, P.; and Schikora, F. 2023. First Time around: Local Conditions and Multi-Dimensional Integration of Refugees. *Journal of Urban Economics*, 137: 103588.

Åslund, O.; and Rooth, D.-O. 2007. Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings. *The Economic Journal*, 117(518): 422–448.

Baba, L.; Schmandt, M.; Tielkes, C.; and Weinhardt, F. 2024. Evaluation der Wohnsitzregelung nach § 12a AufenthG. https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Beitragsreihe/beitrag-band-13-evaluation-wohnsitzregelung.html?nn=283560. Accessed: 2025-05-16.

BAMF. 2024. Initial Distribution of Asylum Seekers (EASY). https://www.BAMF.de/EN/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilung-node.html. Accessed: 2025-05-16.

BAMF-NAvI. 2025. BAMF-NAvI - Integrationskurse. https://bamf-navi.bamf.de/de/Themen/Integrationskurse/. Accessed: 2025-05-16.

Bansak, K.; Ferwerda, J.; Hainmueller, J.; Dillon, A.; Hangartner, D.; Lawrence, D.; and Weinstein, J. 2018a. Improving Refugee Integration through Data-Driven Algorithmic Assignment. *Science*, 359(6373): 325–329.

Bansak, K.; Ferwerda, J.; Hainmueller, J.; Dillon, A.; Hangartner, D.; Lawrence, D.; and Weinstein, J. 2018b. Supplementary Material: Improving Refugee Integration through Data-Driven Algorithmic Assignment. *Science*, 359(6373): 325–329.

Battisti, M.; Peri, G.; and Romiti, A. 2022. Dynamic Effects of Co-Ethnic Networks on Immigrants' Economic Success. *The Economic Journal*, 132(641): 58–88.

Bishu, S. G.; and Alkadry, M. G. 2017. A systematic review of the gender pay gap and factors that predict it. *Administration & Society*, 49(1): 65–104.

Bredtmann, J.; and Höckel, L. S. 2024. Die Bedeutung der beruflichen Weiterbildung für die Arbeitsmarktintegration von Geflüchteten. *Perspektiven der Wirtschaftspolitik*, 25(3-4): 258–272.

Brücker, H. 2024. Arbeitsmarktintegration von Geflüchteten: Verbesserte institutionelle Rahmenbedingungen fördern die Erwerbstätigkeit. *IAB*.

Brücker, H.; Kosyakova, Y.; Rother, N.; Zinn, S.; Liebau, E.; Gider, W.; Schwanhäuser, S.; and Siegert, M. 2025. Exploring Integration and Migration Dynamics: The Research Potentials of a Large-Scale Longitudinal Household Study of Refugees in Germany. *European Sociological Review*, jcaf032.

Brücker, H.; Samy, M. E.; Jaschke, P.; and Kosyakova, Y. 2024. Institutionelle Hürden beeinflussen Umfang und Qualität der Erwerbstätigkeit von Geflüchteten. Research Report 12/2024, IAB-Forschungsbericht.

Brücker, H.; Rother, N. R.; and Schupp, J. 2017. IAB-BAMF-SOEP Befragung von Geflüchteten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. *IAB Forschungsbericht*, 13.

Bürmann, M.; and Tsolak, D. 2025. Much to Lose, No Credentials to Prove It – Educational Aspirations and Intentions of Adult Refugees as Means of Occupational Status Re-Attainment. *European Sociological Review*, jcaf010.

Carrera-Rivera, A.; Ochoa, W.; Larrinaga, F.; and Lasa, G. 2022. How-to Conduct a Systematic Literature Review: A Quick Guide for Computer Science Research. *MethodsX*, 9: 101895.

Castles, S.; Korac, M.; Vasta, E.; and Vertovec, S. 2003. *Integration: Mapping the Field*. Home Office. Research, Development and Statistics Directorate. ISBN 978-1-84473-063-6.

DESTATIS. 2025a. Arbeitslose, Arbeitslosenquoten, Gemeldete Arbeitsstellen: Bundesländer, Jahre. https://www-genesis.destatis.de/datenbank/online/statistic/13211/table/13211-0007. Accessed: 2025-05-02.

DESTATIS. 2025b. Ausländer: Bundesländer, Stichtag, Geschlecht/Altersjahre/ Familienstand, Ländergruppierungen/Staatsangehörigkeit. https://www-genesis.destatis.de/datenbank/online/statistic/12521/table/12521-0021/. Accessed: 2025-05-02.

DESTATIS. 2025c. Bevölkerungsdichte: Bundesländer, Stichtag. https://www-genesis.destatis.de/datenbank/online/statistic/12411/table/. Accessed: 2025-05-02.

DIW. 2024. IAB-BAMF-SOEP-Befragung Geflüchteter 2022, Daten der Jahre 2016-2022.

ESS. 2025. ESS Data Portal. https://ess.sikt.no/en/series/321b06ad-1b98-4b7d-93ad-ca8a24e8788a. Accessed: 2025-05-02.

European Commission. 2024. Pact on Migration and Asylum. https://home-affairs.ec.europa.eu/policies/migration-and-asylum/pact-migration-and-asylum_en. Accessed: 2025-03-21.

European Commission. 2025. Integration - European Commission. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/integration_en. Accessed: 2025-05-16.

Friedman, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5): 1189–1232. Publisher: Institute of Mathematical Statistics.

Gambaro, L.; Neidhöfer, G.; and Spiess, C. K. 2021. The Effect of Early Childhood Education and Care Services on the Integration of Refugee Families. *Labour Economics*, 72: 102053.

Garnett, R. 2023. *Bayesian optimization*. Cambridge University Press.

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

Hahn, E.; Richter, D.; Schupp, J.; and Back, M. D. 2019. Predictors of Refugee Adjustment: The Importance of Cognitive Skills and Personality. *Collabra: Psychology*, 5(1): 23.

Hannafi, C.; and Marouani, M. A. 2023. Social Integration of Syrian Refugees and Their Intention to Stay in Germany. *Journal of Population Economics*, 36(2): 581–607.

Hans, S. 2024. Do Egalitarian Attitudes Promote Integration? In Wonneberger, A.; Stelzig, S.; Weidtmann, K.; and Lölsdorf, D., eds., *Values and Value Change in the Post-Migrant Society*, 25–55. Wiesbaden: Springer Fachmedien. ISBN 978-3-658-45107-3.

Harder, N.; Figueroa, L.; Gillum, R. M.; Hangartner, D.; Laitin, D. D.; and Hainmueller, J. 2018. Multidimensional Measure of Immigrant Integration. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45): 11483–11488.

Herrmann, N. A.; and Rose, M. E. 2025. sprynger: Scriptable bibliometrics using a Python interface to Springer Nature. *SoftwareX*, 31: 102186.

Immigration Policy Lab. 2025. GeoMatch. https://immigrationlab.org/geomatch/. Accessed: 2025-05-07.

INKAR. 2025. INKAR - BBSR. https://www.inkar.de/. Accessed: 2025-05-02.

Jacobs, A. Z.; and Wallach, H. 2021. Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 375–385. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8309-7.

Jaime, S.; and Kern, C. 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, 237–253. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Jones, W.; and Teytelboym, A. 2018. The Local Refugee Match: Aligning Refugees' Preferences with the Capacities and Priorities of Localities. *Journal of Refugee Studies*, 31(2): 152–178.

Kanas, A.; and Kosyakova, Y. 2023. Greater Local Supply of Language Courses Improves Refugees' Labor Market Integration. *European Societies*, 25(1): 1–36.

Khalil, S.; and Tjaden, J. 2025. How Reception Centers Affect the Integration of Asylum Seekers and Recognized Refugees. *European Societies*, 1–32.

Kosyakova, Y.; and Brenzel, H. 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): 123–159.

Kosyakova, Y.; Gundacker, L.; Salikutluk, Z.; and Trübswetter, P. 2021. Arbeitsmarktintegration in Deutschland: Geflüchtete Frauen müssen viele Hindernisse überwinden. Research Report 08/2021, IAB-Kurzbericht.

Kosyakova, Y.; and Kogan, I. 2024. Chapter 19: The Role of Social Capital in Immigrants' and Refugees' Labour Market Integration: Evidence from Germany. chapter Handbook on Inequality and Social Capital. ISBN 978-1-80220-237-3.

Kosyakova, Y.; Salikutluk, Z.; and Hartmann, J. 2023. Gender Employment Gap at Arrival and Its Dynamics: The Case of Refugees in Germany. *Research in Social Stratification and Mobility*, 87: 100842.

Kosyakova, Y.; and Sirries, S. 2017. Large-Scale Immigration and Labour Market Integration: First Lessons from the Recent Past in Germany. *Intereconomics*, 52(5): 263–269.

Makhlouf, K.; Zhioua, S.; and Palamidessi, C. 2021. On the Applicability of Machine Learning Fairness Notions. *ACM SIGKDD Explorations Newsletter*, 23(1): 14–23.

Martén, L.; Hainmueller, J.; and Hangartner, D. 2019. Ethnic Networks Can Foster the Economic Integration of Refugees. *Proceedings of the National Academy of Sciences*, 116(33): 16280–16285.

Mendola, D.; Parroco, A. M.; and Li Donni, P. 2024. We Made It to Germany … and Now? Interdependent Risks of Vulnerability for Refugees in a High-Income Country. *Journal of Ethnic and Migration Studies*, 50(4): 1059–1079.

Meyer, F. 2025a. Integrating Young Refugees into VET: Do German Active Labor Market Programs Make a Difference? *Zeitschrift für Erziehungswissenschaft*.

Meyer, F. 2025b. Refugee Women's Transition to VET in Germany: Examining the Role of Gender Norms and Human Capital Endowments. *Social Inclusion*, 13(0).

Meyer, F.; and Winkler, O. 2023. Place of Residence Does Matter for Educational Integration: The Relevance of Spatial Contexts for Refugees' Transition to VET in Germany. *Social Sciences*, 12(3): 120.

Niculescu-Mizil, A.; and Caruana, R. 2005. Obtaining Calibrated Probabilities from Boosting. In *UAI*, volume 5, 413–20.

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

Ortlieb, R.; Baumgartner, P.; Palinkas, M.; Eggenhofer-Rehart, P.; and Ressi, E. 2024. Employment outcomes of refugee women and men: multiple gender gaps and the importance of high-skill jobs. *Journal of Ethnic and Migration Studies*, 50(20): 4987–5008.

Page, M. J.; Moher, D.; Bossuyt, P. M.; Boutron, I.; Hoffmann, T. C.; Mulrow, C. D.; Shamseer, L.; Tetzlaff, J. M.; Akl, E. A.; Brennan, S. E.; Chou, R.; Glanville, J.; Grimshaw, J. M.; Hróbjartsson, A.; Lalu, M. M.; Li, T.; Loder, E. W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L. A.; Stewart, L. A.; Thomas, J.; Tricco, A. C.; Welch, V. A.; Whiting, P.; and McKenzie, J. E. 2021. PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. *BMJ*, 372: n160.

Phillimore, J. 2021. Refugee-Integration-Opportunity Structures: Shifting the Focus From Refugees to Context. *Journal of Refugee Studies*, 34(2): 1946–1966.

PKS. 2025. Polizeiliche Kriminalstatistik. https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/pks_node.html. Accessed: 2026-08-06.

RCUSA. 2024. Resettlement Process - Refugee Council USA. https://rcusa.org/resources/resettlement-process/. Accessed: 2024-11-05.

Rose, M. E.; and Kitchin, J. R. 2019. Pybliometrics: Scriptable Bibliometrics Using a Python Interface to Scopus. *SoftwareX*, 10: 100263.

Salikutluk, Z.; and Menke, K. 2021. Gendered Integration? How Recently Arrived Male and Female Refugees Fare on the German Labour Market. *Journal of Family Research*, 33(2): 284–321.

Schilling, P.; and Stillman, S. 2024. The Impact of Natives' Attitudes on Refugee Integration. *Labour Economics*, 87: 102465.

Shmueli, G. 2010. To Explain or to Predict? *Statistical Science*, 25(3): 289–310.

Simson, J.; Pfisterer, F.; and Kern, C. 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*, FAccT '24, 1305–1320. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Statistik Portal. 2025. BIP/BWS | Statistikportal.de. https://www.statistikportal.de/de/ugrdl/ergebnisse/wirtschaft-und-bevoelkerung/bipbws. Accessed: 2025-05-02.

Strasser Ceballos, C.; and Kern, C. 2025. Location matching on shaky grounds: Re-evaluating algorithms for refugee allocation. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 2180–2199.

Strazzeri, M. 2021. Assessing the Role of Asylum Policies in Refugees' Labor Market Integration: The Case of Protection Statuses in the German Asylum System. Kiel, Hamburg: ZBW - Leibniz Information Centre for Economics.

Swiss Refugee Council. 2024. Short Overview of the Reception System. https://asylumineurope.org/reports/country/switzerland/reception-conditions/short-overview-of-the-reception-system-2/. Accessed: 2025-05-16.

Tjaden, J.; and Spörlein, C. 2023. How Much Do "Local Policies" Matter for Refugee Integration? An Analytical Model and Evidence from a Highly Decentralized Country. *International Migration Review*, 01979183231205561.

Tsolak, D.; and Bürmann, M. 2023. Making the Match: The Importance of Local Labor Markets for the Employment Prospects of Refugees. *Social Sciences*, 12(6): 339.

UNHCR. 2024a. Projected Global Resettlement Needs 2025. Technical report.

UNHCR. 2024b. Refugees. https://www.unhcr.org/about-unhcr/who-we-protect/refugees. Accessed: 2024-11-11.

United Nations. 1999. Standard Country or Area Codes for Statistical Use. https://unstats.un.org/unsd/methodology/m49/. Accessed: 2025-03-21.

Zschirnt, E.; and Ruedin, D. 2016. Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7): 1115–1134.