A Simulation Framework for Studying the Social Impacts of Algorithm-Based Refugee Matching

CHRISTOPH KERN, LMU Munich, Germany
JAKOB KAPPENBERGER, University of Mannheim, Germany
FREDERIC GERDON, University of Mannheim, Germany
CLARA STRASSER CEBALLOS, LMU Munich, Germany
DARIA SZAFRAN, University of Mannheim, Germany
FLORIAN RUPP, Mannheim University of Applied Sciences, Germany
RUBEN BACH, University of Mannheim, Germany

The integration chances of refugees in their host country are critically shaped by the contextual conditions of the location to which they are assigned upon arrival. Several research groups have developed algorithmic tools to optimize refugee-location matching, with the overall aim of improving refugees' integration outcomes. These tools are used in a highly sensitive context and thus their design, social impacts, and potential long-term consequences need to be systematically assessed. To investigate such effects, we propose an agent-based simulation framework that allows to simulate different allocation mechanisms and to study their impacts on integration outcomes. We illustrate the simulation framework in the German context by comparing the current approach of the *Königsteiner Schlüssel* (i.e., quasi-random allocation) with the algorithm-based procedure *GeoMatch*. We study each procedures' impacts on both labor market and social integration and assess structural effects on inequalities between subgroups of asylum seekers. The decision models and agents' characteristics are based on the IAB-BAMF-SOEP survey of asylum seekers and refugees in Germany. Our study shows how agent-based models can be used to study unintended consequences of algorithmic allocations of asylum seekers in dynamically changing social environments.

Keywords: algorithmic matching, agent-based model, integration, refugees, fairness

Reference Format:
Christoph Kern, Jakob Kappenberger, Frederic Gerdon, Clara Strasser Ceballos, Daria Szafran, Florian Rupp, and Ruben Bach. 2025. A Simulation Framework for Studying the Social Impacts of Algorithm-Based Refugee Matching. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 5 pages.

1 Introduction

Motivation. The integration of asylum seekers into the labor market is a central topic of political and scientific debates [8]. Successful labor market integration depends on a multitude of intertwined factors and becomes more

Authors' Contact Information: Christoph Kern, LMU Munich, Munich, Germany, christoph.kern@lmu.de; Jakob Kappenberger, University of Mannheim, Mannheim, Germany; Frederic Gerdon, University of Mannheim, Mannheim, Germany; Clara Strasser Ceballos, LMU Munich, Munich, Germany; Daria Szafran, University of Mannheim, Mannheim, Germany; Florian Rupp, Mannheim University of Applied Sciences, Mannheim, Germany; Ruben Bach, University of Mannheim, Mannheim, Germany.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

likely if there is a match between asylum seekers' characteristics and the characteristics of their resettlement location [2, 3, 9, 11]. However, current governmental procedures that assign asylum seekers to locations typically do not consider the match of asylum seekers with host location characteristics. For instance, in Germany and Switzerland, the allocation process can be considered (quasi-)random. In Germany, the allocation follows the *Königsteiner Schlüssel*, which assigns asylum seekers to federal states based on the population size and tax revenue of the target state [5].

Against this background, several research projects have proposed algorithm-based matching tools that consider asylum seekers' or refugees' characteristics for resettlement decisions, with the aim to improve their integration outcomes (e.g., [1, 6]). While these tools directly impact vulnerable populations, systematic evaluations or forecasts of their potential long-term impacts are scarce. Such assessments are complicated by the fact that location matching tools present a special type of algorithmic decision-making (ADM) system with unique properties: (1) there are no inherently correct "ground truth" location assignments against which the algorithms' recommendations can directly be evaluated, and (2) prediction-based assignment algorithms include both a prediction and a matching step which jointly affect the allocation decision. This challenges the application of key concepts and auditing tools from the field of algorithmic fairness and, at the same time, elevates the need to assess the complex downstream effects and potential feedback loops that may arise when algorithmic matching systems are deployed in dynamic social environments.

Algorithmic location matching. Prediction-based matching tools such as *GeoMatch* [6] and *Annie™ Moore* [1] assign refugees to locations based on predictions of integration outcomes. In short, these 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 then generated for newly arriving refugees. These refugees are assigned to locations that maximize an optimality criterion (e.g., average employment) subject to constraints (e.g., location capacity). For *GeoMatch*, it has been shown that algorithmic matching can achieve overall relative *predicted* employment gains of approximately 41% for the U.S. and 73% for Switzerland [6]. While optimizing alternative integration outcomes is, in principle, possible, to the best of our knowledge, *GeoMatch* has not been implemented and evaluated for any other outcome than labor market integration.

Assessing the implications of design decisions of algorithmic matching tools is critical. For example, economic integration – the single target optimized by current *GeoMatch* applications [6] – is just one among many dimensions of integration. To understand the implications of algorithmic matching more broadly, it is necessary to evaluate how optimizing for labor market integration may affect outcomes in other dimensions of integration. For instance, allocation decisions could correlate with and affect the demographic composition of local populations. This can in turn affect the chances of social integration, here understood as the level of social interaction between asylum seekers and natives [8] as well as non-natives in the host society [2]. Thus, feedback loops may result from how algorithm-based matching shapes the demographic composition of resettlement locations, affecting integration chances of future asylum seekers. Such dynamic impacts cannot be evaluated by simply aggregating predicted matching outcomes, but rather need scenario-based assessments of how the deployment of algorithmic matching tools optimized for specific outcomes may interact with regional conditions over time. Finally, raising issues of fairness, there may be subgroups among asylum seekers for which the models work worse than for others, even when compared to random allocation.

Contribution. Given these critical issues with algorithm-based allocations of asylum seekers, we propose an empirically informed agent-based simulation and present its application with data from the IAB-BAMF-SOEP survey of asylum seekers and refugees in Germany [7]. With this model, we simulate the allocation procedures based on asylum seekers' and location characteristics. Agent-based simulations allow to (1) assess how optimizing for solely one integration outcome (labor market integration) may lead to unintended consequences by impacting other integration outcomes (social integration), (2) investigate feedback loops resulting from changes in local demographic compositions induced by matching and (3) study how the use of matching procedures can reinforce social inequalities, i.e., have disparate impacts between subgroups of asylum seekers.

With this research, we contribute to the growing literature on using algorithmic decision-making systems in sensitive social contexts such as asylum seeker allocation. By drawing on a simulation setup, we highlight dynamic interaction mechanisms that constitute potential challenges and can lead to unintended negative consequences. More generally, this study contributes to research on computational frameworks for evaluating the social impacts of algorithmic decision-making in high-stakes, dynamic environments. These insights are also relevant from a policy-making perspective as they may inform critical stakeholder discussions on whether and how to apply such decision-making approaches.

2 Agent-based simulation model

Objective. We present an agent-based modeling framework that simulates two allocation mechanisms for asylum seekers in Germany: quasi-random assignment based on the *Königsteiner Schlüssel*, which governs this process in reality, and algorithmic matching inspired by the *GeoMatch* approach. We compare how these approaches perform with respect to both labor market and social integration outcomes. Furthermore, our simulation setup can reveal possible temporal dynamics in macro-level demographic compositions, resulting from how allocation processes shape the environments for future agents.

Data and models. The environment of the agent-based model represents Germany's 16 federal states. We build two machine learning models to predict the two integration dimensions relevant to our approach, namely labor market and social integration, and use them both for the prediction of outcomes in our simulation and to devise an allocation strategy similar to *GeoMatch*. For this purpose, we draw on data from the IAB-BAMF-SOEP survey of asylum seekers and refugees in Germany, integrated into the Socio-Economic Panel (SOEP) study since 2016. Random samples from the Central Register of Foreigners are drawn to gather information about asylum seekers and refugees who immigrated to Germany since 2013 [7]. Additionally to the survey data, we supplement the models with data on the unemployment rates as well as the community sizes of residents with foreign nationalities in the different federal states, respectively. For the dimension of labor market integration, our model estimates the probability of a given individual finding work in the respective federal state (employed vs. not employed), which we derive from a GPBoost model, combining gradient boosting and mixed effects to incorporate the panel structure of the survey data [10]. For social integration, we create an index of two items measuring the frequency of interactions between an asylum seeker and native and non-native individuals, following [2], which we then convert to a binary variable (frequent vs. non-frequent interactions) and again predict using GPBoost.

Simulation setup. The agents in the simulation are asylum seekers who, after their arrival and initial registration, are allocated to one of the federal states. All agents in the model own a series of attributes derived from the survey

data (e.g., age, country of origin, gender, education, German language skills). Every time step in the simulation corresponds to one year, starting in 2013 and finishing in 2021 (the period for which we have data from the IAB-BAMF-SOEP survey).

Each iteration of the simulation first involves creating new agents. Throughout the simulated time frame, the volume of agents spawned is proportional to the number of asylum seekers recorded in Germany for that specific year. Moreover, the subpopulation of each year is randomly drawn from data of respondents who arrived in said year. The newly spawned agents are then distributed according to one of the two allocation schemes, and the current iteration is concluded. Following allocation, we compute the integration outcomes of agents in the federal state they were assigned to. After three time steps (i.e., years) in the simulation, agents are removed as the mandatory domicile obligation ("Wohnsitzauflage") of asylum seekers in Germany ends [4].

3 Preliminary results

Our preliminary experiments over 25 simulations per allocation strategy indicate that deploying algorithm-based allocation targeted at maximizing the labor market integration of asylum seekers indeed results in a significantly higher share of working agents in our simulation. As Figure 1 shows, this is achieved by maximizing the outcomes in a series of federal states while accepting lower employment rates in other states when compared to the Königstein baseline. Thus, the average improvement comes at a cost, as the algorithmic allocation scheme concentrates high-potential agents in states where the model predicts the greatest success. Regional variation in social integration differences are less pronounced but negative impacts for specific states still stand out. These findings demonstrate the negative side-effects of algorithmic location matching of asylum seekers. Building on these preliminary results, we will investigate feedback loops induced by the allocation process and analyze how matching procedures may introduce disparate outcomes across different subgroups of refugees. We further highlight the need to jointly consider multiple integration objectives in algorithmic matching tools.

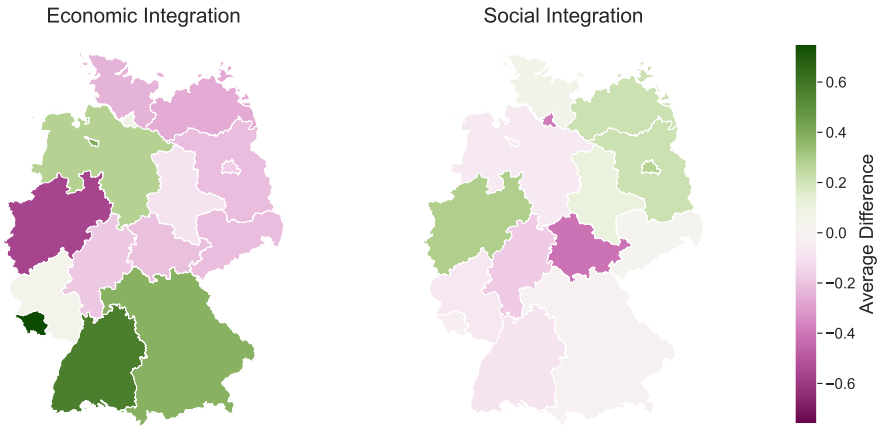

Fig. 1. Average relative differences regarding employment (left) and social integration (right) between simulated outcomes after algorithm-based allocation versus Königsteiner Schlüssel. Green coloring marks states in which algorithm-based allocation performed better than the baseline (and vice versa).

References

[1] 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

[2] 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

[3] Daniel Auer. 2018. Language roulette–the effect of random placement on refugees' labour market integration. *Journal of Ethnic and Migration Studies* 44, 3 (2018), 341–362.

[4] 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

[5] BAMF. 2024. *Initial Distribution of Asylum Seekers (EASY)*. https://www.BAMF.de/EN/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilung-node.html

[6] 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

[7] DIW. 2023. IAB-BAMF-SOEP-Befragung Geflüchteter 2021, Daten der Jahre 2016-2021. https://doi.org/10.5684/soep.iab-bamf-soep-mig.2021

[8] 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* 115, 45 (2018), 11483–11488. https://doi.org/10.1073/pnas.1808793115

[9] Tobias Müller, Pia Pannatier, and Martina Viarengo. 2023. Labor market integration, local conditions and inequalities: Evidence from refugees in Switzerland. *World Development* 170 (2023), 106288.

[10] Fabio Sigrist. 2023. Latent Gaussian Model Boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 2 (Feb. 2023), 1894–1905. https://doi.org/10.1109/TPAMI.2022.3168152

[11] Louisa Vogiazides and Hernan Mondani. 2020. A geographical path to integration? Exploring the interplay between regional context and labour market integration among refugees in Sweden. *Journal of Ethnic and Migration Studies* 46, 1 (2020), 23–45.