Ethos: Designing Algorithmic Mechanisms for Increased Fairness, Stakeholder Empowerment, and Systemic Accountability

Liya Mulugeta
Niloufar Salehi, Ed.
David A. Wagner, Ed.

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-294
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-294.html

December 1, 2023

Ethos: Designing Algorithmic Mechanisms for Increased Fairness, Stakeholder Empowerment, and Systemic Accountability

by

Liya Kirubel Mulugeta

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Wagner, Chair
Professor Niloufar Salehi
Professor Victoria Robinson

Spring 2023

Abstract

Ethos: Designing Algorithmic Mechanisms for Increased Fairness, Stakeholder Empowerment, and Systemic Accountability

by

Liya Kirubel Mulugeta

Master of Science in Electrical Engineering and Computer Science

University of California, Berkeley

Professor David Wagner, Chair

A key concern of policymakers who use student-assignment algorithms is increasing diversity and fairness. Drawing on two years of research into the student assignment algorithmic mechanism of Ethiopia, I studied how the collective values of stakeholders can shape the parameters of algorithmic mechanisms. I found that prioritizing the collective values of stakeholders in the initial stages of the algorithm design process enhances university diversity, empowers students to submit truthful rank-order lists, and establishes accountability mechanisms for universities. One of the main contributions in this thesis is a novel assessment of mechanism fairness, as defined by having: (1) an absence of **justified envy**, (2) an absence of **lack of information**, and (3) an absence of **misalignment of values**. Ethos, the core technical contribution in this thesis, consists of a machine learning-backed acceptance rate quiz for Ethiopian public universities, an informative portal for Ethiopian students, and a generalized student-university matching algorithm. My process for designing and evaluating Ethos consists of two in-person user studies (Study 1, n=33; Study 2, n=40) in which I identified and assessed the real-world impact of my system and algorithm parameters. I argue that listening to and prioritizing the collective values of stakeholders is critical to building diverse and fair algorithmic mechanisms and offer generalizable methods for doing so.

To my extremely loving and supportive parents, Kirubel and Rozina.

Acknowledgement

My deepest gratitude belongs to my family, for giving me unparalleled love and encouragement. Thank you, Dad, for accompanying me through countless hours of interviews with the students in Ethiopia, always listening to me with a patient ear, and giving me wise advice during my Bachelor's and Master's. Thank you, Mom, for encouraging me to pursue my Master's, always being there for me, and giving me an abundance of love during my Bachelor's and Master's. My parents were my rock throughout this process. I also want to thank to my sister, Saron, and my brother, Kalab, for believing in me and encouraging me. Finally, I am tremendously grateful for the other wonderful individuals who have made my academic journey possible and transformative.

- Thank you to Professor Niloufar Salehi for being a caring and incredible advisor throughout my candidacy. She has been a thoughtful and welcoming professor whom I highly respect and admire for introducing me to the HCI research community and the field of participatory algorithm design. I simply would not be the researcher I am today without her incredible expertise, kindness, and life-changing mentorship. I will forever be grateful for the opportunity to have been one of her graduate research students.

- Thank you to Professor Victoria Robinson, who taught me how to view any research that I carry out through a critical theory lens. She taught me what it truly means to critically study and analyze the world around me as a student of ethnic studies. Thank you, Victoria, for changing how I view the world.

- Thank you to Professor Jay Chen, who taught me how to carry out research within the limiting context of resource-constrained communities.

- Thank you to Audrey Sillers, for showing me unconditional kindness and support. Thank you, Audrey, for being the heart of my support system on campus. Thank you to the rest of the Electrical Engineering and Computer Science department for believing in me and very generously giving me all of the support that I needed to bring it to life.

- Thank you to all of the students and staff at the schools who gave me their valuable time and honest feedback during the studies, including at: Addis-Coder, Debre Berhan Secondary School, Medhanilem Secondary School, St. Mary's Secondary School, St. Joseph's Secondary School, and Addis Ababa University.

- Thank you to the stakeholders at the Ethiopian Ministry of Science and Higher Education who gave me all the resources and data that I needed to run my models and test my algorithms, including but not limited to: Dr. Zelalem Assefa, Dr. Solomon Benor, Yonatan, and Eyob.
- Finally, thank you to Ryan Mullins, my mentor at Google Research, for providing me with his extremely valuable advice early on.

Contents

Contents **iv**

List of Figures **vi**

List of Tables **viii**

1 Introduction **1**
- 1.1 The Problem . . . 1
- 1.2 Goals . . . 2
- 1.3 Contributions . . . 3
- 1.4 Thesis Outline . . . 3

2 Background **5**
- 2.1 Student Assignment Algorithms . . . 5
- 2.2 The MSHE's Student Assignment Algorithm . . . 6
- 2.3 Mechanism Fairness . . . 8
- 2.4 Summary . . . 10

3 Prior Work **11**
- 3.1 Algorithmic Mechanisms and Theories of Justice . . . 11
- 3.2 Participatory Design and User Empowerment . . . 12
- 3.3 Participation and Fairness in Algorithm Design . . . 13
- 3.4 Agent-Empowerment in Algorithm Design . . . 15
- 3.5 Summary . . . 17

4 Study 1: Need Finding **18**
- 4.1 Methods . . . 18
- 4.2 Results . . . 21
- 4.3 Discussion . . . 28

5 Ethos: Multi-Criteria Student Assignment Mechanism **30**
5.1 The Informative Portal 30
5.2 The Machine-Learning Model 30
5.3 The Matching Algorithm 34

6 Study 2: System Evaluation in Needs-Based Setting **39**
6.1 Methods 39
6.2 Results 42
6.3 Discussion 42

7 Future Study: System Evaluation in Assets-Based Setting **44**
7.1 Methods 45
7.2 Discussion 47

8 Discussion **49**
8.1 Recommendations 50
8.2 Limitations 50
8.3 Positionality and Reflexivity 51

9 Conclusion **52**

Bibliography **54**

List of Figures

2.1 Model of the Current Centralized University Matching Process in Ethiopia. The possible rank-order lists of the 43 public universities are represented by the n-ordered tuples $\alpha = (u_1, u_2, ..., u_n) \in \mathbb{N}^{43}$ (the set of all 43-ordered tuples of natural numbers). 7

4.1 Card-sorting. 20
4.2 Model of Student Circles of Influence on their Rank-Order List of 43 Public Universities. 21

5.1 Homepage of Ethiopian Public Universities Portal with AI Assistant. . . 31
5.2 (a) Anonymized Testimonials from Students in Study 1 (in Dark mode and with the dimensions of a tablet). (b) Directory of all the universities on portal. 31
5.3 (a) Hawassa University Page (in English). (b) Hawassa University Page (Translated by Google to Amharic); there are accuracy limitations to this translation feature, as an off-the-shelf plugin. 32
5.4 (a) All Departments for Hawassa University. (b) Demographics and Categorical Rankings for Hawassa University. 33
5.5 The percentage of students who got assigned to their respective ranked choice. For example, a little less than 40% of students got assigned to their first-choice in their given rank-order list. Going back to our analysis in Chapter 4, students do not know how to rank universities beyond their top three. Moreover, as we can see above, the MSHE's matching algorithm prioritizes matching students to their top ten universities. On top of my analysis in Chapter 4, this finding was one of the several reasons why I chose to specify the top ten universities as the ones which students can edit for the ML-backed quiz. 34
5.6 Our predicted acceptance rate model achieved a high accuracy rate for each university, ranging from 90% to 99%. 35

5.7 The model-generated recommended rank-order list for the given student data. 36
5.8 The predicted acceptance rates to each university for the current student based on the current algorithm and the proposed algorithm. 37

6.1 Ethos is a system for algorithmic decision making that was designed using a participatory algorithmic design framework. This framework translates the collective values of stakeholders into the parameters of the algorithm. I present this framework as a way to increase fairness, stakeholder empowerment, and systemic accountability in algorithmic decision-making. 40

7.1 The user flow detailing how end users can interact with the bot service. 48
7.2 The high-fidelity wireframe of the bot service. 48

List of Tables

4.1 The subcategories, descriptions, statement examples, and resulting algorithm parameter I for the theme of **Infrastructure and Internet**. . . . 23
4.2 The subcategories, descriptions, statement examples, and resulting algorithm parameter D for the theme of **Available Departments**. 24
4.3 The subcategories, descriptions, statement examples, and resulting algorithm parameter Q for the theme of **Quality of Education**. 25
4.4 The subcategories, descriptions, statement examples, and resulting algorithm parameter C for the theme of **Campus Environment**. 26
4.5 The subcategories, descriptions, statement examples, and resulting algorithm parameter U for the theme of **University Resources**. 27

Chapter 1

Introduction

1.1 The Problem

Algorithmic mechanisms are increasingly being adopted within decision-making across various public sector domains such as healthcare, child welfare, and law enforcement [68, 12, 16, 61, 32, 66]. These systems impact people's lives and have faced numerous criticisms from researchers over their discriminatory practices [14, 15] and unintended consequences to vulnerable communities [13, 48]. There is a growing concern that systems which are designed without consideration to the needs and values of the affected communities are in danger of harming them [22]. A growing body of literature on participatory algorithm design implements social choice theory as a framework for these systems to collect and aggregate individual stakeholders' preferences [7, 18, 27, 28, 30, 41, 42, 46]. However, the path from gathering and aggregating individual stakeholder preferences to providing an algorithmic mechanism that can be practically implemented is not straight-forward [50, 49, 54, 70, 71]. In this thesis, I employ a bottom-up approach to design an algorithmic mechanism whose parameters were shaped by the collective values of stakeholders.

In several countries, including Chile, Turkey, Germany, Taiwan, and the U.K., a high-stakes application of algorithmic mechanisms is the national assignment of students to public universities through a centralized university matching process [2, 52, 64]. In this system, students are assigned to a university among a limited number of universities. A student may also be assigned to zero universities.

1.2 Goals

The goal of my work was to research the student assignment algorithmic mechanism in Ethiopia with the aim of developing a novel algorithmic mechanism which increases the diversity of its public universities' student bodies. This case study offers a practical algorithmic mechanism that is rooted in value-sensitive design and emphasizes user empowerment by directly using the collective values of the stakeholders to set the values of the algorithm which will facilitate their allocation. Ethos, the core technical contribution of this work, consists of a machine learning-backed acceptance rate quiz for Ethiopian public universities, an informative portal for Ethiopian students, and a generalized student-university matching algorithm.

My process for designing and evaluating Ethos consists of two in-person user studies (a need-finding study and a system evaluation study) with a total of 73 participants. I cover my need-finding study (Study 1) in Chapter 4 and my system evaluation study (Study 2) in Chapter 6.

The purpose of Study 1 was to identify the collective needs, values, and aspirations of students who undergo the national student assignment algorithmic mechanism in Ethiopia. For this study, I conducted 33 semi-structured interviews with students across public and private secondary schools in Ethiopia during Winter 2022.

The purpose of Study 2 was to evaluate the diversity and fairness outcomes of my proposed algorithmic mechanism. For this study, I conducted a within-subjects needs-based study using a randomized controlled trial in Summer 2023. I recruited 40 students from various regions and secondary schools in Ethiopia as my sample. This study deployed an informative portal and my proposed student-university algorithmic mechanism that can be generalized to other contexts and countries. My combination of qualitative and quantitative methods across both of these studies allowed me to triangulate my research so that I could have a stronger understanding of the student assignment algorithmic mechanism for Ethiopia. This research puts algorithm research, critical theory, and human-centered computing into conversation with each other.

1.3 Contributions

The core of the contributions made by this thesis is the Ethos system, outline in Chapter 5, which consists of a novel assessment of mechanism fairness and generalizable student-university matching algorithm, a machine learning-backed acceptance rate quiz for Ethiopian public universities, and an informative portal for Ethiopian students in Chapter 5. I also provide two in-person user studies in Chapter 4 and 6 (Study 1, n=33; Study 2, n=40) in which I identify and isolate the variables of my algorithm by answering the following research questions:

- **RQ1:** How can we translate the student's specific needs, aspirations, and values into parameters of a matching algorithm to increase the fairness of the algorithmic mechanism?
- **RQ2:** How does exposing the student to their personalized acceptance rates to the 43 public universities as well as a recommended rank-order list of the 43 public universities affect the fairness of the matching algorithm?

1.4 Thesis Outline

The remainder of my thesis is organized into chapters as follows:

- **Chapter 2** provides background information and related work on the key topics discussed in my thesis, including the current centralized matching process in Ethiopia, classical literature on student assignment algorithms and mechanism design, and my proposed novel assessment of mechanism fairness.
- **Chapter 3** describes groundwork on theories of justice in algorithmic mechanisms, participatory design, and user-empowerment in algorithm design.
- **Chapter 4** discusses the experimental design and results analysis of Study 1. This study informs the basis for my proposed matching algorithm in Chapter 5. From this study I can define the following three main findings which will serve as the backbone of the rest of my research:
    1. The majority of students rank the same public university as their first choice.

2. Typically, students randomly rank the majority of the 43 public universities.
3. The lack of transparency from both the Ministry and the universities forces students to rely on word-of-mouth and social signaling from their Circles of Influence to inform their rank-order list of the 43 public universities.

- **Chapter 5** details my process for developing the Ethos system, which consists of my informative portal, my machine-learning backed quiz, and my proposed matching algorithm.
- **Chapter 6** presents the experimental design and results analysis of Study 2, where I test the Ethos system's performance in a needs-based setting.
- **Chapter 7** outlines the experimental design of a future study where one can test the Ethos system's performance in an assets-based setting.
- **Chapter 8** covers my work's recommendations for future work, some of the limitations in my studies, and my positionality in this research project.
- **Chapter 9** offers concluding remarks. Specifically, I summarize the key contributions of my thesis.

Chapter 2

Background

In this chapter, I will provide an overview of related work in student placement algorithms and background information relevant to the goals and scope of my thesis. First, I will discuss the classical literature of student assignment algorithms. Second, I will provide an overview of the Ethiopian Ministry of Science and Higher Education's (MSHE) student assignment algorithm. Third, I will define the fairness property of mechanisms and how my research contributes to our understanding of this property.

2.1 Student Assignment Algorithms

Existing literature in student assignment algorithms builds upon two core matching mechanisms [2, 53, 56], Gale and Shapley's Deffered Acceptance algorithm [31] and Gale's Top Trading Cycles (TTC) [62]. One of the most frequently-used real-life student assignment algorithms is the Boston Student Placement algorithm [2, 45]. This algorithm and its minor modifications are presently implemented in various cities including Boston, Seattle, Minneapolis, Lee County, and Florida [2]. This algorithmic mechanism begins by considering, for each school, all the students who listed that school as their first choice and allocating them by their priority order one at a time until there are no available seats left or there are no students who have listed it as their first choice [44]. Next, consider the subset of students who have not been placed in any school in the previous step. For each school with available seats, only those students who have listed the school as their second choice are considered. The priority order of students within each school is again followed to place these students one at a time until either no seats are left or there are no students remaining who have listed the school as their second choice. This cycle repeats until all students are placed.

The Boston algorithm has limitations [44]. Specifically, the algorithm is not strategy-proof, as it can give students incentives to misrepresent their true preferences in order to maintain their priority for certain schools. The algorithm's design forces students to think strategically and make non-truthful submissions, which can lead to suboptimal outcomes. In addition, school district authorities often advise students and their parents to make strategic choices [33]. Empirical evidence has confirmed the strategic behavior of students under the Boston algorithm. Chen and Sonmez conducted an experiment and found that 80% of the subjects chose to misrepresent their preferences under the algorithm [20]. This is due to the fact that students fear losing their priority for certain schools and believe that strategic choices will increase their chances of being assigned to their preferred school. Misstating preferences is so prevalent that even suggestions encouraging such behavior have appeared in the press [26]. The high incidence of strategic behavior in the Boston algorithm raises concerns about this mechanism's fairness.

2.2 The MSHE's Student Assignment Algorithm

In student assignment algorithmic mechanisms, each university has a certain number of available seats. For each university, there is a strict priority order of all the students, and each student has a strict order of preferences for all the universities [31]. The final student-university allocations are allocated by a clearinghouse through a classical student assignment algorithm that is rooted in mechanism design [55, 4, 62]. These clearinghouses differ from country to country, from the Ministry of Education in one country to a private organization in another country [44]. In the U.S., for example, the placement of medical students to residency options is determined by a clearinghouse called the National Residency Match Program [60]. Additionally, many cities in the U.S., such as San Francisco and Boston, also delegate clearinghouses to assign students to primary and secondary schools [2, 45, 58]. In recent years, these classical student assignment algorithms have been researched and modified with the hope that the newer versions will increase diversity within schools and optimize the outcomes of all of the stakeholders [45, 8, 34, 53, 43, 37, 1]. However, in practice, these algorithms have several limitations and instead decreased diversity in schools [58, 35].

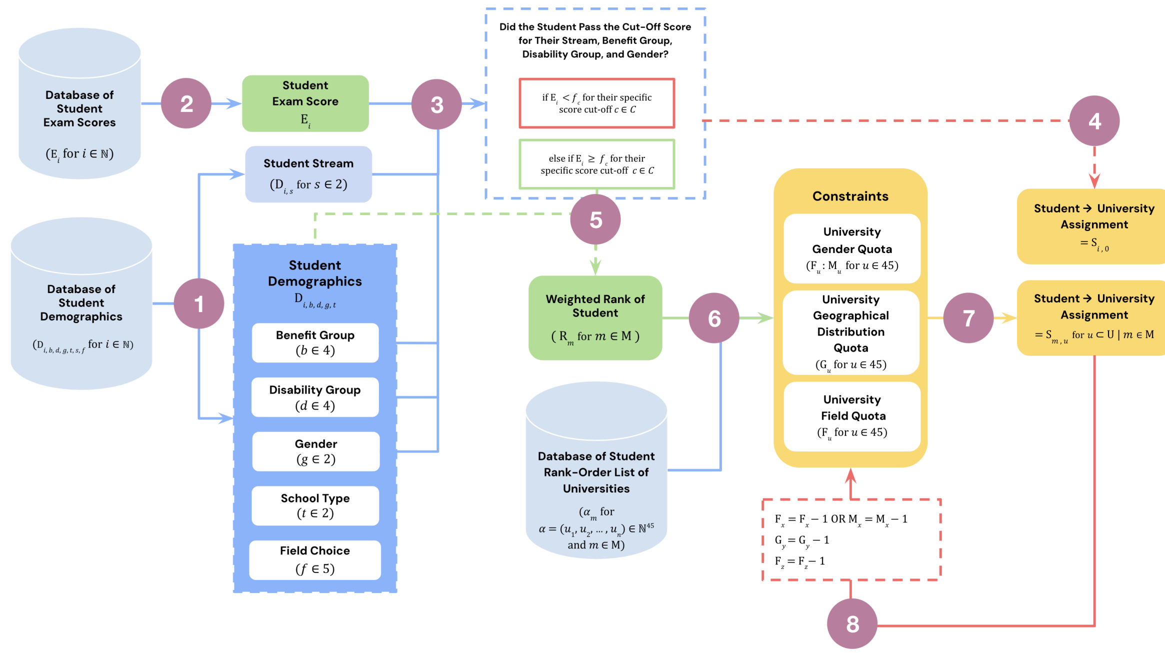


Figure 2.1: Model of the Current Centralized University Matching Process in Ethiopia. The possible rank-order lists of the 43 public universities are represented by the n-ordered tuples $\alpha = (u_1, u_2, ..., u_n) \in \mathbb{N}^{43}$ (the set of all 43-ordered tuples of natural numbers).

The clearinghouse for Ethiopia's national student assignment mechanism is the Ministry of Science and Higher Education in Ethiopia (MSHE), which has a set of criteria to admit and place students in higher institutions, including student entrance exam minimum requirements and university quota constraints. I developed a high-level system model (see Figure 2) to outline the input parameters, constraints, and output (set of student-university assignments) according to a specification document provided by the MSHE[1]. From this high-level system diagram and data, I were able to create a machine learning model which can predict the acceptance rate of a given student to any university in the current centralized matching process with a high-level of accuracy (over 90%).

[1]The MSHE had legal and privacy protections which limited them from sharing the code behind the current matching algorithm with me, but shared other details and data about the process.

2.3 Mechanism Fairness

Mechanism fairness determines how the preferences of the students are aggregated and how the matching is performed. Traditionally, mechanism fairness is based on the absence of justified envy [44, 3]. A student has justified envy for another student assigned to a university if they prefer the university to the one that they are assigned to and they have higher priority for it than at least one student who is placed to it [44]. In my work, I define a novel assessment of mechanism fairness as having: (1) an absence of **justified envy**, (2) an absence of **lack of information**, and (3) an absence of **misalignment of values**. The second property arises from Study 1, where I found that most students randomly ranked 90% of the universities on their rank-order list, in large part due to a lack of information on them. The third property is also informed from my analysis in Study 1, where I found that the university qualities need to align with the needs, desires, and aspirations of the students.

I analyze the impact of this novel definition of fairness for the case study of student assignment algorithmic mechanisms during Study 2. In this study, I deploy my proposed system, Ethos. Ethos consists of three main technical components: 1) an information portal, 2) a machine learning-based quiz for suggesting universities, 3) a multi-criteria matching algorithm. The machine learning model used in Ethos can predict a given student's acceptance rates to each university based on the MSHE's current student-assignment algorithm. I built this model on over 160,000 data points of previous student-university assignments. To develop it, I used one-hot encoding and max-voting on an ensemble of three different machine-learning models.

The ethos of this work is that building more equitable societal mechanisms requires critically examining our technology's systemic impacts and its positionality within a community in supporting or suppressing user agency. In this work, I worked to respect this concept by listening to the students on their drawbacks, desires, and aspirations for the student assignment algorithmic mechanism in Study 1 and using these results to inform the production of my algorithmic mechanism in Study 2. I showcase how the normative gap in the current student assignment algorithmic mechanism is that it does not align with the actual demands for justice for more diverse student bodies at the Ethiopian public universities. This is because it neglects to consider the student's Circles of Influences which might deter them from submitting truthful and accurate rank-order lists of the universities. Within the context of mechanism design, which is the backbone of algorithmic mechanisms such as student-university assignment, this normative gap disrupts the classical definition of mechanism fairness. Ultimately, I argue that expanding our sociotechnical imaginary requires avenues that give agency to users in shaping algorithmic systems which affect them.

Mechanism **fairness** determines how the preferences of the students are aggregated and how the matching is performed. However, the notion of fairness can be subject to varying interpretations. In my work, I define a novel assessment of mechanism fairness that is outlined below as having:

- The absence of **justified envy**. Justified envy is a classical definition of fairness in mechanism design. A student has **justified envy** for another student assigned to a university school if he prefers the university to the one that he is assigned to and he has higher priority for it than at least one student who is placed to it (Kesten 2004).
- An absence of **lack of information** on the university. I base this definition on my finding from Study 1 that most students randomly rank the universities due to lack of information.
- The absence of **misalignment of values** between the university qualities and the needs, desires, and aspirations of the student. This property is also informed from my analysis in Study 1.

Therefore, I define allocation as **fair** if there is no student-university assignment such that the student has either justified envy, a lack of information, or a misalignment of values.

A well-known fact in mechanism design is the incompatibility between fairness and Pareto efficiency [44]. That is, there is no mechanism that is both Pareto-efficient and fair. This result is known as the impossibility theorem, which was first proven by economist Kenneth Arrow in his seminal paper "Social Choice and Individual Values" in 1951 [6]. Arrow's theorem states that it is impossible to design a voting mechanism that satisfies a set of reasonable criteria, including Pareto-efficiency and fairness, without violating at least one of them. If a Pareto efficient and fair allocation exists for a given student placement problem, then it is the one selected by the student optimal stable mechanism. In addition, A Pareto efficient and fair allocation may not always exist and if it exists, it is unique [60]. By extension, Kesten points out that the Student Optimal Stable

Mechanism is not Pareto efficient and the Top Trading cycles Mechanism is not fair [44].

2.4 Summary

In this chapter, I outline the classical literature on student assignment algorithms and mechanism design. Next, I discuss the MSHE's student matching algorithm. Finally, I motivate the need for a novel assessment of fairness in algorithmic mechanisms.

Chapter 3

Prior Work

In this chapter, I will present the connection between mechanism design and a theory of justice. I will leverage this connection to motivate the need for challenging the ecosystem that a matching algorithm lives in and encourage matching algorithm designers to work backwards from the narratives of the agents to ensure that we develop the most equitable matchings. Next, I will discuss the previous literature on transparency, participation, and fairness in algorithm design. Finally, I will discuss approaches for designing algorithms which incorporate user empowerment, resist techno-determinism, and are conscious for communities that are resource-constrained or vulnerable.

3.1 Algorithmic Mechanisms and Theories of Justice

In "Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms," the authors came up with four design implications for student assignment systems, including: providing relevant and accessible information, aligning and realigning algorithmic objectives with community goals in mind, reconsidering how stakeholders express their needs and constraints, and making appropriate,reliable avenues for recourse available [58]. One of the main takeaways of this paper was that student assignment algorithms exist within and to uphold a political ideology that privileges individual choice sometimes at the cost of other values, such as democracy, resource equality, and desegregation.

This takeaway aligns with Hitzig's proposal that mechanism design enacts a theory of justice [38]. Hizig bases this proposal analysis on two unusual features of the Boston algorithm redesign: 1) that the economic theory is enacted in the school system and 2) that it draws on an elaborate, but unarticulated, normative framework. Putting these two features together suggests that mechanism design can be reframed through an ideal theory of distributive justice. From this, she argues that there is a normative gap between the implicit normative theory of the mechanism and the actual demands of justice.

3.2 Participatory Design and User Empowerment

There has been growing interest in incorporating distributive justice and individual preferences into algorithmic decision-making through the use of participatory algorithm design [72, 16, 50, 73, 74, 21]. Our research relies heavily on the framing which participatory design provides. Incorporating this framework into our research is necessary for evaluating how students are choosing to participate in this school choice mechanism.

Toyama's "amplification thesis," posits that technology's only impact is amplification [67]. Therefore, a major consideration should be: what is amplified? Decision-making systems that are built and operated without input from the community they affect are in danger of harming them [22]. The concept of technology giving agency back to its users is well-articulated in Chambers' work, where he argues that researchers must respect the basic human right of vulnerable communities to conduct their own analysis and listen to their inputs as we research a solution for that community [19]. In this work, we respect this concept by listening to the students on their drawbacks, desires, and aspirations for the centralized university matching process in Study 1. Additionally, Burrell's Material Eco-Systemic Approach defines an improved ethic of design in ICT where ICT methods should account for, support, and amplify the agency of its users [17]. By extension, we can deduce that the true value of technology in resource-constrained settings is how much it can empower and liberate its users, such as in the case study of Fisher's KickStart [29]. In order to achieve this end goal, our work employs a four-step research method: (1) Ethnographic Research, (2) Need-Finding Research, (3) Iterate and Refine Prototypes, and (4) Onsite User Testing [40, 39]. The first two steps are covered in Study 1 while the final two steps are covered in Study 2. Our goal with this paper is to illustrate how to practically develop a connection between the algorithm developer and the algorithm agents, in order to create a feedback loop that encourages algorithmic accountability and agent empowerment.

3.3 Participation and Fairness in Algorithm Design

Participation

A number of emerging methods for participatory algorithm design have proposed collecting and aggregating individual stakeholder preferences to create algorithmic systems that represent the values and goals of those stakeholders. There is greater importance in incorporating distributive justice and individual preferences into algorithmic decision-making through the use of participatory algorithm design. My research relies heavily on the framing which participatory design provides. Incorporating this framework into my research is necessary for evaluating how students are choosing to participate in this school choice mechanism.

In "Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms", the authors came up with four design implications for student assignment systems, including: providing relevant and accessible information, aligning and realigning algorithmic objectives with community goals in mind, reconsidering how stakeholders express their needs and constraints, and making appropriate, reliable avenues for recourse available [58]. One of the main takeaways of this paper was that student assignment algorithms exist within and to uphold a political ideology that privileges individual choice sometimes at the cost of other values, such as democracy, resource equality, and desegregation. This takeaway aligns with Hitzig's proposal that mechanism design enacts a theory of justice [38].

In order to come to this conclusion, the authors analyze over a dozen interviews between families and people who have supported families in navigating the school choice and enrollment process. In their results, they state that one of the families' critical pain points was that the information provided about schools did not account for the difficult trade-offs or relevant context that families must navigate when choosing between schools. In order to address this concern, among others such as the limits of informational resources for promoting educational equity, they advocate an assets-based design approach to enrollment support and prioritizing the value of personalized support and trusting relationships to delivering relevant and helpful information. In my own research, I incorporate their proposal for an assets-based design approach in Study 3.

The authors of "Stakeholder Participation in AI: Beyond Add Diverse Stakeholders and Stir", outline approaches which stakeholders could take in order to increase participatory design [23]. They do this by analyzing the different past approaches to increasing participation in design and then introducing a new framework which they define as the five “dimensions of participation”. This multidimensional framework includes a series of five questions:

1. Why is participation needed?
2. What is on the table?
3. Which stakeholders should be involved?
4. What form does their participation take?
5. How is power distributed among the participating stakeholders and between stakeholders and technology designers/engineers?

One of the reasons this paper is powerful is because it evaluates sociotechnical power dynamics and the current method of empowering stakeholder agency along the axis of meaningful participation in computational algorithms. Similarly, it forces the reader to consider a need for a more democratic approach to designing algorithms because of how current AI development practices typically prioritize stakeholder’s preferences over deliberative democratic decisions on the forms of participation for that algorithm.

At the same time, there are also negative consequences to exclusively using preferences as the sole method of participation and a greater need to consider a broader range of values and needs, especially those of vulnerable groups such as those with accessibility difficulties or limited resources [58]. It is important to note that value-sensitive design does not provide an explicit ethical theory to designate what kinds of values should be supported. Therefore, in addition to an understanding of implicit values and politics, their analysis includes a commitment to justice and accepting refusal as a legitimate way of engaging with technology. In my research, I make it a point to tackle these concerns behind traditional rank-order preference lists, especially for those with limited resources or accessibility difficulties, in Study 3. I also attempt increase my incorporation of a participatory framework in my studies by providing an avenue for students to have transparent insight into and provide feedback on the current student assignment algorithm and my proposed student assignment algorithm during Study 2.

Fairness

In "Lost in Translation: Reimagining the Machine Learning Life Cycle in Education", the authors outline the danger of employing inequitable computational algorithms in order to increase justice in education [51]. The authors found that there are two undermined yet critical algorithmic design considerations which computer scientists must consider when they are developing machine learning algorithms in the space of educational justice: Translating Education Goals via Problem Formulation and Translating Predictions to Interventions.

One of the most important nuances of this paper was the fact that it was built on the critical expertise of scholars in educational research alongside the author's own expert knowledge in machine learning research to form research findings which truly aim at closing translational gaps between the two domains. A specific research finding which stood out as being closely related to my work while pulling on principles from the concepts of value-sensitive design and design justice was in section 4.3, which outlined why computer scientists must design inputs of the algorithm with education equity in mind.

It is also imperative that we acknowledge the current audits of algorithmic tools and the criticisms of the fairness framework. The current audits of algorithmic tools serve to protect against certain algorithmic harms which Bandy outlines in "Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits" such as discrimination, distortion, exploitation, and misjudgement [9]. As for criticisms of the fairness framework, the authors of "Fairness and abstraction in sociotechnical systems" emphasize how any fair machine learning solution requires a commitment to either learning new social science research skills or partnering with social scientists in order to mitigate the chances of the technical solution from falling into the five traps which they outline as solutionism, the ripple effect, formalism, portability, and framing.

3.4 Agent-Empowerment in Algorithm Design

In this section, I will explore how my project attempts to respectfully consider the community that my proposed algorithmic mechanism will effect by employing ICT research methods. In other words, these research methods will help ensure that my algorithm is built in community with the locals.

ICT Research Methods in Algorithmic Mechanism Design

The concept of techno-determinism is viewing technology as an autonomous agent which disrupts our ability to see technology and its design as something that might be shaped by humans and their political and ethical concerns. Techno-determinism is a fundamentally inadequate theory of the relationship between technology and society. Rather than dealing with the underlying conditions, technical solutions which do not systemically improve the lives of their users are usually short-term responses to the fundamental issue they were attempting to solve. The problem with techno-determinism is that it is a euphemism for reformist ideologies in the form of code. Moreover, in Toyama's "amplification thesis", we learn how technology's only impact is amplification [67]. This line is stating how we must develop technology products that are not only fulfilling user needs but aligned with the axis of the local community values to provide maximum value to the end-users.

The concept of technology giving agency back to its users is echoed in Chambers' paper, where he argues that researchers must respect the basic human right for poor people to conduct their own analysis and listen to their inputs as we research a solution for that community [19]. In my project, I respect this concept by listening to the students on their drawbacks, desires, and aspirations for the centralized university matching process in Study 1. Additionally, Burrell's Material Eco-Systemic Approach defines an improved ethic of design in ICT where ICT methods should account for, support, and amplify the agency of its users [17].

By extension of Burrell's approach, we can deduce that the true value of technology in resource-constrained settings is how much it can empower and liberate its users, such as in the case study of Fisher's KickStart [29]. In order to achieve this end goal, my project employs a four step research method: (1) Ethnographic Research, (2) Need-Finding Research, (3) Iterate and Refine Prototypes, and (4) Onsite User Testing [39]. Namely, my research completes the first two steps of this ICT research method in Study 1 and the last two steps across Study 2 and Study 3.

Algorithm Research Being in Community with Locals

Building on the amplification thesis, society-impacting algorithms that operate in isolation from the community they affect are in danger of harming them. They must operate in symphony with the society which they impact. My goal with this thesis is to develop a connection between the algorithm designers and the algorithm agents, in order to create a symbiotic relationship that encourages algorithm accountability through feedback loops, redistributes the power between the technology and the algorithm agents, and ultimately, gives agency back to the agents.

Algorithmic researchers doing projects, like this one, for resource-constrained communities, do not have the luxury to ignore the ethics of our algorithms. If we want to re-imagine the community's future with our algorithms, we must be critically reflective and start by questioning the reality of the affects and systems which the current algorithms have on the communities we are trying to uplift.

At the same time, we cannot be seduced by the appeal of an algorithm which fixes one visible problem but perpetuates many hidden systemic problems. In some societies, there are intrinsic and local hierarchies that cause value gaps, which can limit the design of the algorithm to protect the agents, such as by "design(ing) within the patriarchy" in order to protect women [65]. In cases such as these, it is clear that algorithms must be cautiously innovative and integrate its solutions intentionally, respectfully, and slowly into those local communities in order to have their technology accepted by locals.

3.5 Summary

In this chapter, I present the connection between algorithmic mechanism design and theories of justice. Next, I cover groundwork in participatory and user empowerment methods of algorithm design. Then, I share some of the limitations and precautions we must take when designing algorithms within resource-constrained or vulnerable communities. Finally, I motivate the need for applying research methods from Information Communication Technologies (ICTs).

Chapter 4

Study 1: Need Finding

In Study 1, I conducted ethnographic field research and need-finding interviews to understand user needs and desires from the student assignment algorithmic mechanism. Given the novelty of the phenomenon, I employed a similar approach to Smyth et al. and avoided a hypothesis-oriented method of analysis, opting instead for inductive reasoning to identify key themes [63]. This approach is standard among several well-known qualitative analysis techniques.

RQ1: How can we translate the student's specific needs, aspirations, and values into parameters of a matching algorithm to increase the fairness of the algorithmic mechanism?

4.1 Methods

My goal in this Institutional Review Board (IRB) approved research study was to understand the values and needs of the high school students when submitting their rank-order list of the 43 public universities to the Ethiopian Ministry of Science and Higher Education for the centralized university matching process. Within the context of algorithm design, I aimed to understand these values and needs so that I could convert them into variables that would be measured within my proposed algorithm. To achieve this goal, I conducted an ethnographic action research study that employs ICT research methodologies.

Data Collection

I began the study by conducting a set of formal, in-person, and semi-structured interviews (n=33). Data collection took place over a 4-week period in Addis Ababa, Ethiopia in the winter of 2022. In this study, 27 of the 33 participants were 12th-grade secondary school students. The remaining 6 participants were public university students. Of the 33 participants, 18 identified themselves as female and 15 identified themselves as male. 17 of the students went to a public secondary school while the other 10 students went to a private secondary school. Both of these user subject groups were interviewed on the values and priorities which they believed would impact (or had impacted, if they were university students) their ranking of the 43 universities during the centralized university matching process.

It was important for me to collect data that represented a diverse range of socioeconomic levels and a balanced gender distribution in order to have the closest possible representation of the millions of 12th grade students across the nation that undergo the student-university algorithmic mechanism in Ethiopia.

Data Analysis

To analyze the main themes discussed by the participants during this study, I transcribed data gathered over three weeks of interviews and labeled them accordingly using line-by-line open coding. I revised the labeling through an iterative process, and then used axial coding to extract the relationship between themes. I identified five emerging themes that students wanted information on in order to rank their universities: (1) Infrastructure and Internet, (2) Available Departments, (3) Quality of Education, (4) Campus Environment, and (5) University Resources. I expanded on the subcategories for each of these emerging themes, as well as how they translate to our proposed algorithm parameters, in Table 1, 2, 3, 4, and 5.

The proposed algorithm parameter for each subcategory represents the importance or relevance of that subcategory to the student's decision-making process when selecting a university. Each of these parameters will be assigned a weight and used to inform the proposed matching algorithm. Following my rigorous analysis of the interviews, I developed a conceptual framework to illustrate the Circles of Influence perceived by the typical student participant during the centralized university matching process in Ethiopia. The model

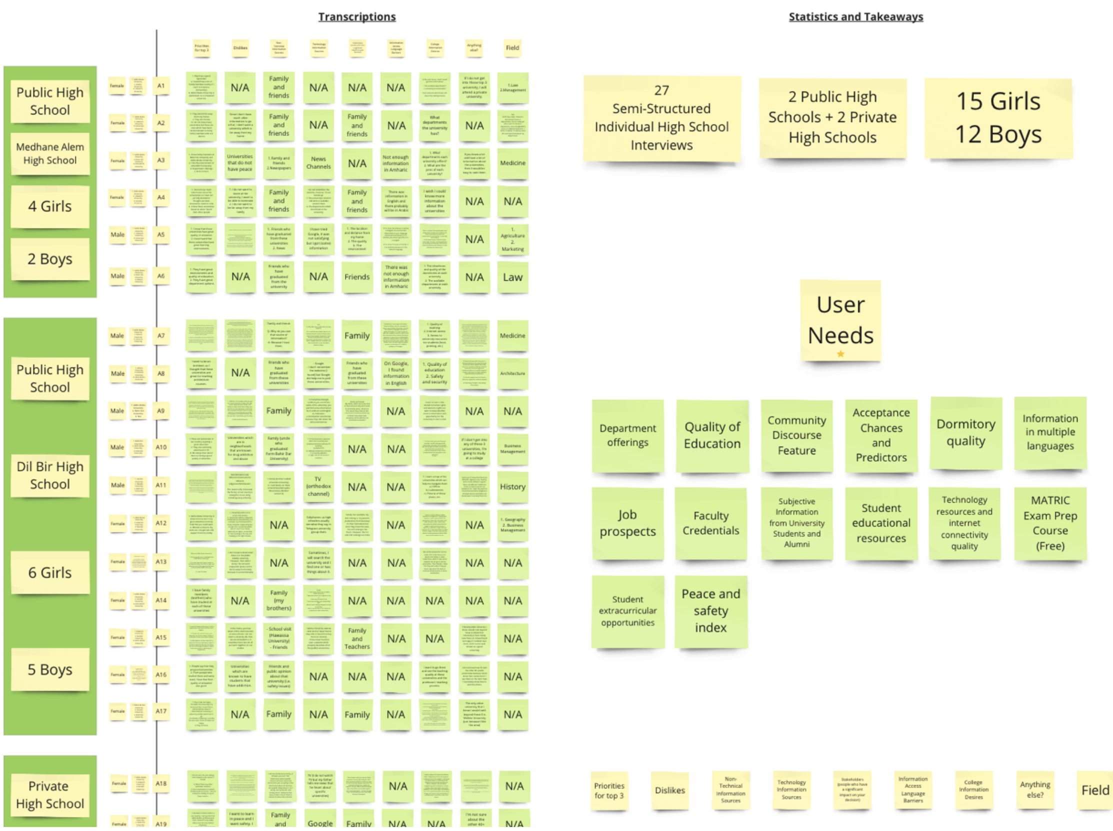


Figure 4.1: Card-sorting.

shown in Figure 4.2 depicts the prevailing sentiment among the student participants. The sentiment that the students shared is that they lack agency in the selection of their assigned university, as they are insufficiently equipped with factual information about the 43 public universities in Ethiopia, which they are required to rank in order of preference. These key takeaways are supplemented by three findings.

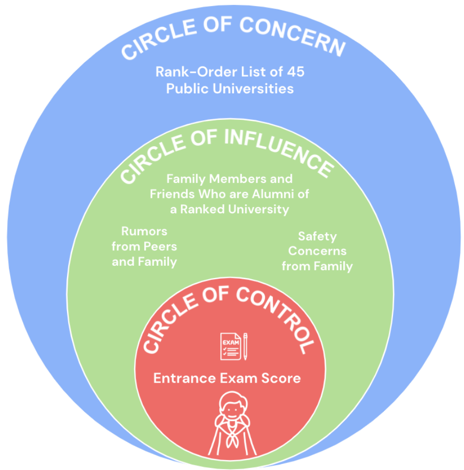


Figure 4.2: Model of Student Circles of Influence on their Rank-Order List of 43 Public Universities.

4.2 Results

Qualitative Analysis

I present three main findings from this study. The first finding is that all of the student interviewees lacked confidence in making a rank-order list of university preferences beyond their top three choices, despite having to rank 43 universities. Hence, many of the students ranked the same well-known university as their first choice.

The next finding was that students were forced to trust and rely on word-of-mouth and social signaling subjective information from alumni with whom they had connections or family members. For example, when choosing which university to list as their first choice, students often deferred their decision to a family member who was an alumnus of a certain public university: “[M]y brother studied there and he told me to choose it because that is where those with top scores go” [A9].

The last finding was the most surprising for me: students did not have a single standardized, reliable, and factual information source to consult in order to learn about the attributes of the 43 universities they had to rank in their preferences for the university matching process. These 43 public universities had minimal to no online presence. Therefore, unless the high school provided an information packet to their students, which was a rare case and typically only mentioned within the private high schools that I visited, students were forced to rely on subjective and limited information from their Circles of Influence to guide the development of their rank-order list. When questioned about their values and priorities that led them to pick their top three universities, several of the students stated that their top three choices were "totally based on rumors. I do not know [the validity] of any of these rumors" [A18]. **In other words, students were randomly ranking the majority of the 43 universities on their rank-order list.** From an algorithmic design perspective, this finding implies that regardless of the algorithm I develop, unless I address this concern, I will not be able to optimize the matches for neither the student nor the university side.

Subcategories	Description	Statement Example	Parameter
Building Infrastructure and Campus Internet Connectivity	The infrastructure of the campus buildings and the internet connectivity.	"I do not want to go to universities without adequate infrastructure and electricity. For example, my friend wanted to study computer engineering but she got into a university which doesn't even have steady electricity for lights, so she had to start working in the day and studying in the night classes."	$I_{infrastructure}$
Dormitory Sanitation and Food Quality	The quality of living facilities, particularly dormitories, and the quality of food options available to the students.	"One of my priorities is the food quality. . . Most people consider this because they are concerned about their health. If we are not offered good food options we will not be healthy so we will not able to study and we may fail our exams..."	I_{dorms}

Table 4.1: The subcategories, descriptions, statement examples, and resulting algorithm parameter *I* for the theme of **Infrastructure and Internet**.

Subcategories	**Description**	**Statement Example**	**Parameter**
Major Offerings	The different academic departments and fields of study offered by a university.	"If someone knows what they want to know beforehand, then they will target universities which teaches what they want. For example, in Ethiopia, Gondar University is said to be good for medicine studies, so it would be good if students who want to study medicine know this kind of information."	D_{fields}
Major Selection Process	The field selection process and acceptance rate at a university.	"I want to study 2 different fields (business management or geography). However, I can't make a choice. For geography, I have a lot of interest and I want to learn it. But, I don't have information on which universities it is taught at, what kind of courses it requires, or whether it aligns with my interests. So, if I got information on these kinds of things, it would really make the decision making process of which field I want to chose easier for me."	$D_{selection}$

Table 4.2: The subcategories, descriptions, statement examples, and resulting algorithm parameter D for the theme of **Available Departments**.

Subcategories	Description	Statement Example	Parameter
Quality of Pedagogy	Refers to the teaching methods, strategies, and the overall quality of education.	"I want to know the universities teaching strategy and quality of education."	$Q_{teaching}$
Practical Learning Opportunities for STEM Majors	The availability and quality of hands-on learning experiences for students studying science, technology, engineering, and mathematics (STEM) fields.	"Engineers should learn 75% of their material through practice and experience. However, at that university, engineering students learn 75% of their material through theory and 25% through experience, therefore, when they come out and get hired, they do not know anything and make mistakes."	$Q_{practice}$
Job Placement Rate	The likelihood of graduates finding employment in their desired field after graduation.	"The second reason why I chose these universities as my top three is because they have great post-graduate opportunities nearby. For example, if I were to go to Addis Ababa University and study properly, I can find a good job or I can become a lecturer there when I graduate. So, in other words, I would have good [job] opportunities."	$Q_{employability}$

Table 4.3: The subcategories, descriptions, statement examples, and resulting algorithm parameter Q for the theme of **Quality of Education**.

Subcategories	Description	Statement Example	Parameter
Peace, Safety, and Security	The peace, safety, and security of the campus environment.	"I will not rank universities that are in conflict zones higher on my list [...] I want to learn in peace and I want safety. I do not want conflict."	C_{safety}
Drug Prevention and Tolerance	The tolerance level for drug use on campus.	"I do not want to go to a university which tolerates drug addiction. At those kinds of universities, you could find gangsters and people like that, and then you don't know where you'll end up."	C_{drugs}
Distance from Part-Time Job Opportunities	The distance between the student's home and the university they are assigned to.	"I do not want to study in university because I want to work in the daytime and do night classes at a college. Even if I get into Addis Ababa University with my marks, I would still work in the day and study at night."	C_{jobs}
Distance from City and Landmarks	The distance of the student to the city and landmarks.	"I want to rank Adama University because I know the area of Adama well, it's not new to me, I can live there and I know the language of that area. And I say Hawassa because I'd love to go out in Hawassa, that's my dream."	$C_{recreation}$

Table 4.4: The subcategories, descriptions, statement examples, and resulting algorithm parameter C for the theme of **Campus Environment**.

Subcategories	**Description**	**Statement Example**	**Parameter**
Student Educational Resources	The availability of academic resources, including campus maps, libraries, laboratories.	"I wish I could see a map of all of the campus buildings and resources for the students like the offices, libraries, and laboratories."	$U_{learning}$
Student Extracurricular Opportunities	The range and quality of extracurricular activities offered by the university, including sports and clubs.	"For Bahir Dar . . . they have great extracurricular options. For example, I like football, and they have great clubs, I would love to play for their football club. So, Bahir Dar's resources, their atmosphere, and their campus' beauty are what draws me to that school."	$U_{extracurriculars}$
Student Mental Health Resources	The available mental health resources, such as counseling services for students struggling with mental health issues.	"I want to learn more about the healthcare and mental health resources at the universities. Because university life might disrupt our mental health, maybe, and it might challenge us because we are apart from our family so it will be important to know about our healthcare options."	U_{health}

Table 4.5: The subcategories, descriptions, statement examples, and resulting algorithm parameter U for the theme of **University Resources**.

Quantitative Analysis

The participants were questioned on their current process for finding information on the universities which they would need to rank. Of all of the secondary school students, 74% of them said they were going to rank Addis Ababa University as their number one choice and 55% said they were going to rank Hawassa University within their top three choices. 14% of the secondary school students were unsure of their top three choices. All of the public university participants, said that they ranked the university which they currently attend, Addis Ababa University, as their number one choice during the centralized university matching process. This statistic aligns with the fact that 74% of the secondary school participants said that they were going to rank Addis Ababa University as their first choice.

Of the 43 universities that they would be asked to rank during the centralized university assignment process, an overwhelming 92% of the secondary school students said that they knew how to rank their top three university choices and that they did not know the names of (or were going to randomly rank) the remaining 42 universities. This percentage includes both private and public secondary school students. This result was shocking as an important assumption of the algorithm behind the centralized university assignment is that the students know how to rank the 43 universities to accurately reflect their preferences. Even for their top three choices, many of the secondary school students stated that these top three choices were purely based on rumors.

4.3 Discussion

In this study, I found that regardless of any matching algorithm process I use, the results suggest that there are underlying systemic issues in the centralized university matching process in Ethiopia that need to be addressed in order for my matching algorithm to operate optimally within the mechanism. I found that the critical issue with the prior algorithmic mechanism was that it falsely assumed that students had agency over the rank-order list of the 43 universities which they submitted to the Ethiopian Ministry of Science and Higher Education. In reality, the majority of students we interviewed felt as if their university matching was in their external locus of control. To address this problem and receive rank-order lists which reflect the truthful and accurate preferences of the student participants, I propose re-structuring the algorithmic mechanism to include providing factual information to the students about the 43 universities. In Study 2, I will explore how to achieve this desired outcome.

Design Implications

From this study I can define the following two design implications:

- **In order for participants to be able to make informed rankings, they need accessible information.**
- **In order for universities to get diverse student bodies, they need to explain to the students how their university adheres to the student's needs, aspirations, and desires.**

Future Work

Following the results of this study, one of my recommendations for future work in algorithm design for vulnerable groups is to take a leap of faith on the agents to assist in the direction of the algorithm design. The dominant themes and takeaways that I extracted in this study would not have been discovered if I had not taken a moment to speak to the users in order to find out why students were not getting matched to the universities which they wanted and why universities were not getting a diverse set of students. To connect this back to my discussion of techno-determinism in Chapter 2.5, as algorithm designers, we are outsiders. Therefore, we must learn how to give the agents a space where we simply listen as they voice their opinions. We should not claim authority over the direction of the algorithm by default. Those who are the most affect by the algorithms must teach us what they have experienced and be the compass of my algorithmic design process.

Chapter 5

Ethos: Multi-Criteria Student Assignment Mechanism

In this chapter, I describe my proposed system, Ethos. Ethos consists of three main technical components: 1) an informative portal, 2) a machine learning-based system for suggesting universities, 3) a multi-criteria matching algorithm.

5.1 The Informative Portal

I built this portal using an industry-standard technical stack and standard data-visualization libraries, including: **React.js**, **TypeScript.js**, **d3.js**, **HTML**, and **SCSS**. This ability of this portal to easily be deployed, scale, and be responsive was crucial for me, because it needed to be able to function regardless of the device a student accessed it on. This necessity was driven by an observation I made in Study 1, where I noticed that students had varying levels of access to technology. In order to operate my study in Chapter 6, I needed my portal to work efficiently and smoothly, even if the student only had a mobile phone.

5.2 The Machine-Learning Model

I developed a machine learning model which can predict a given student's acceptance rates to each university based on the current student-assignment algorithm. To develop this model, I used one-hot encoding, an ensemble of three different models, and max-voting to return the class with the maximum number of votes. The three models I voted on were: a Random Forest Classifier model, a Logistic Regression model, and a Gradient Boosting Classifier model.

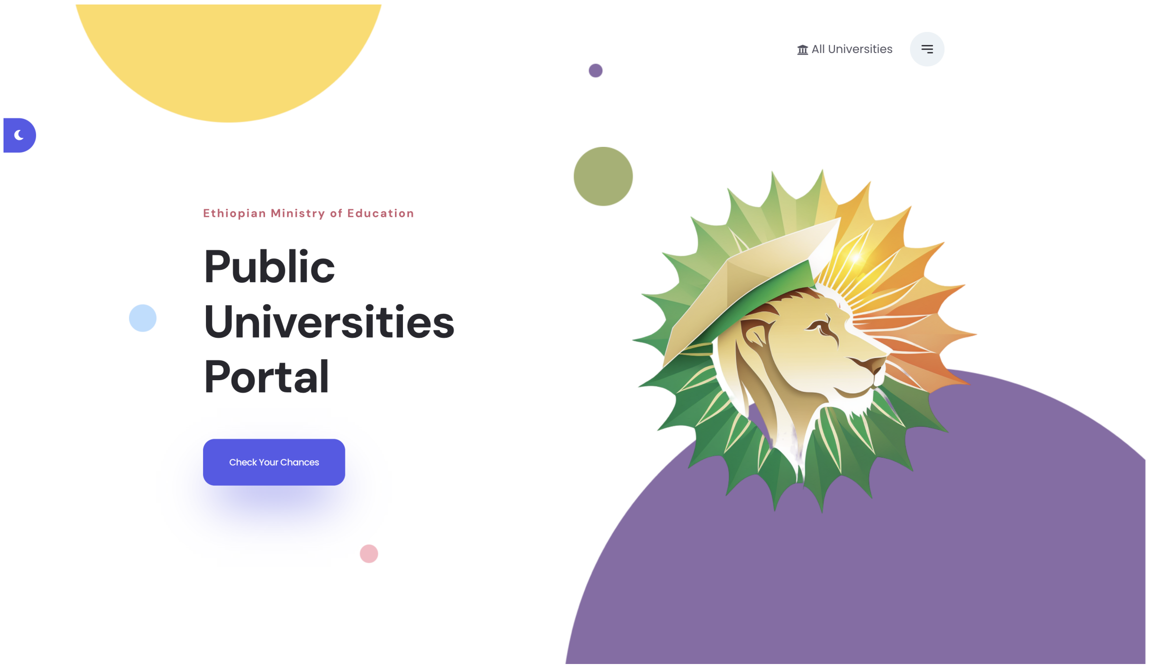
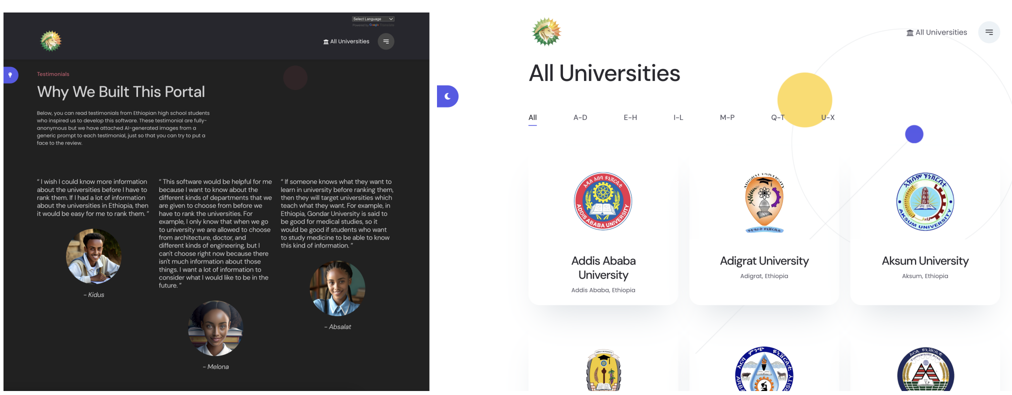
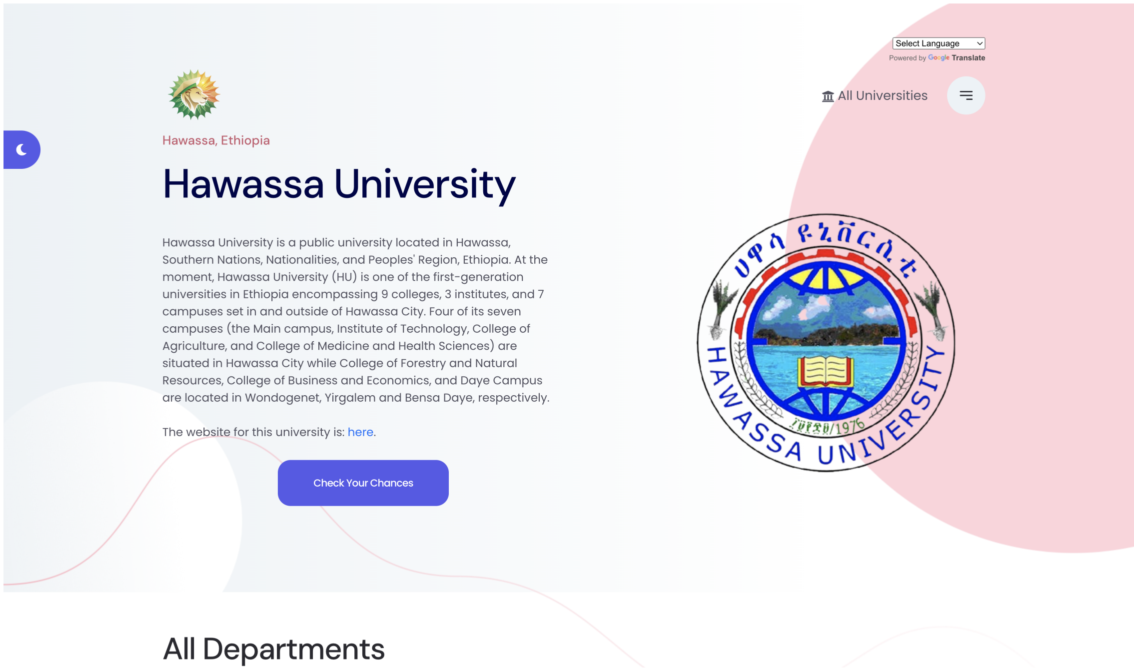


Figure 5.1: Homepage of Ethiopian Public Universities Portal with AI Assistant.

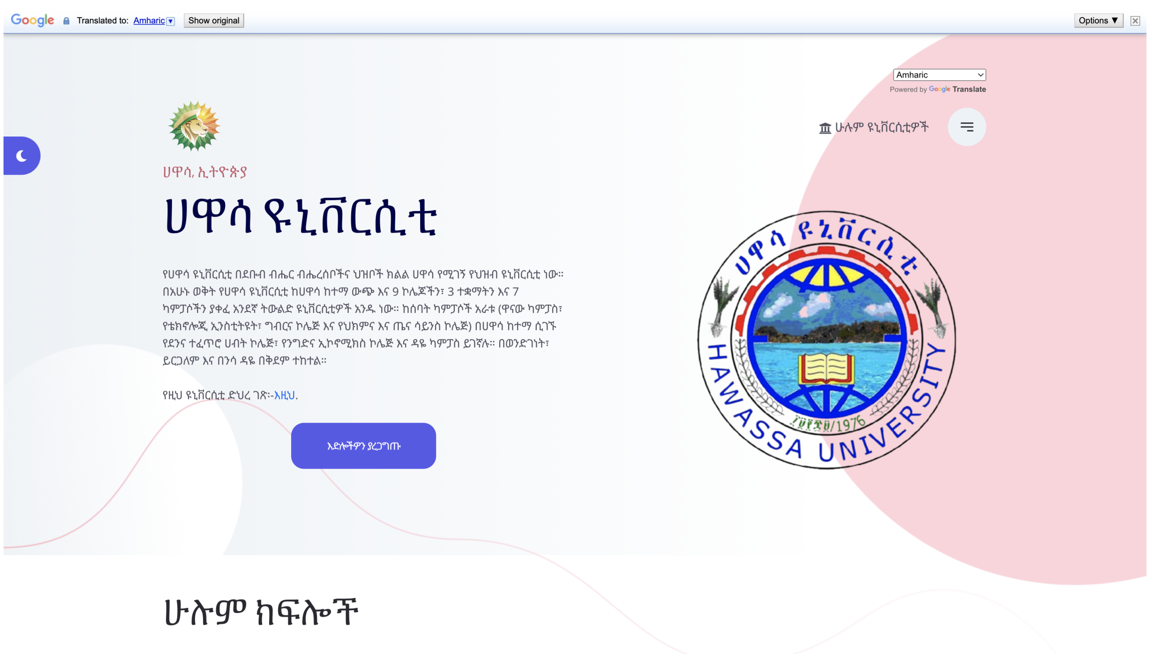


(a) (b)

Figure 5.2: (a) Anonymized Testimonials from Students in Study 1 (in Dark mode and with the dimensions of a tablet). (b) Directory of all the universities on portal.

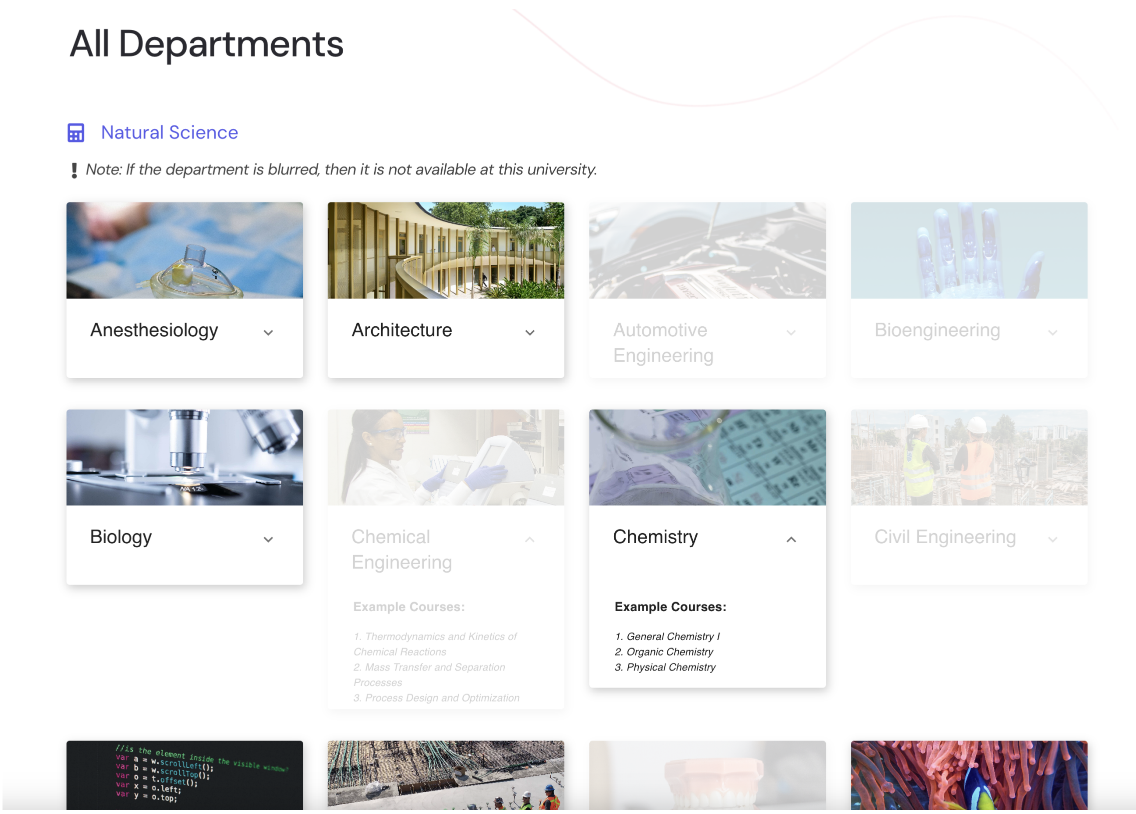


(a)

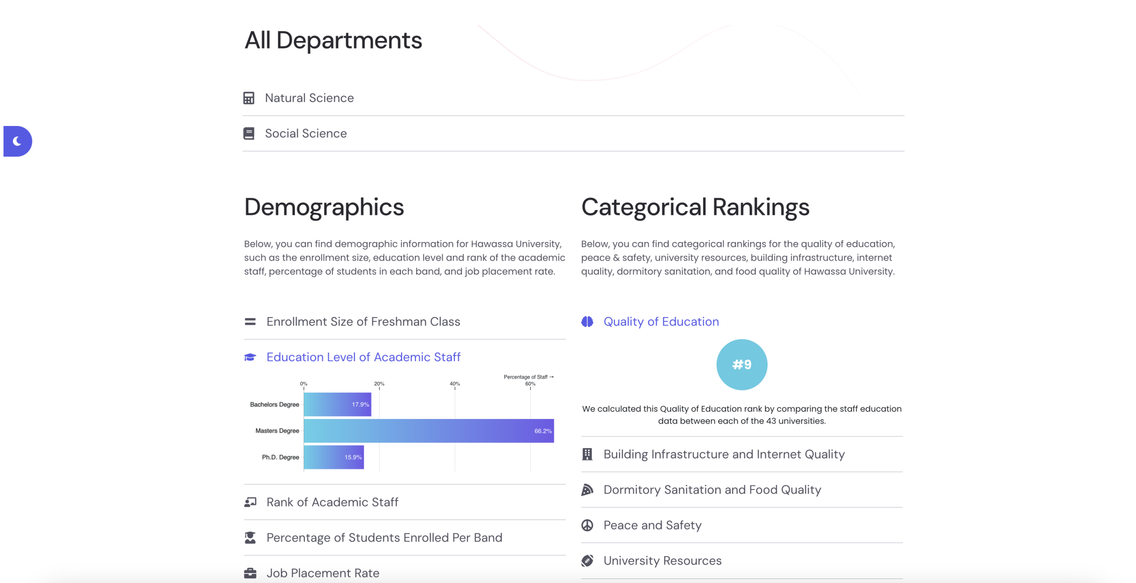


(b)

Figure 5.3: (a) Hawassa University Page (in English). (b) Hawassa University Page (Translated by Google to Amharic); there are accuracy limitations to this translation feature, as an off-the-shelf plugin.



(a)



(b)

Figure 5.4: (a) All Departments for Hawassa University. (b) Demographics and Categorical Rankings for Hawassa University.

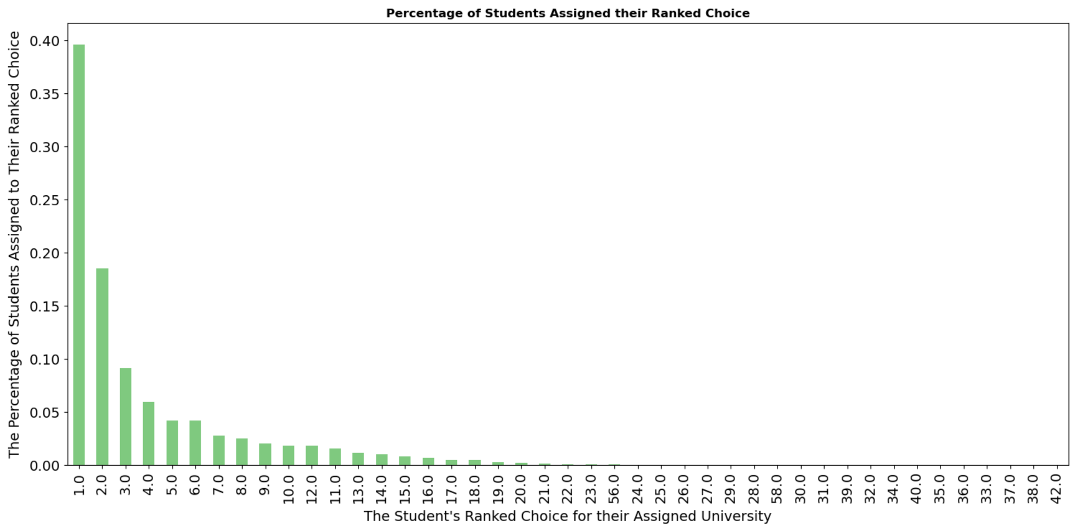


Figure 5.5: The percentage of students who got assigned to their respective ranked choice. For example, a little less than 40% of students got assigned to their first-choice in their given rank-order list. Going back to our analysis in Chapter 4, students do not know how to rank universities beyond their top three. Moreover, as we can see above, the MSHE's matching algorithm prioritizes matching students to their top ten universities. On top of my analysis in Chapter 4, this finding was one of the several reasons why I chose to specify the top ten universities as the ones which students can edit for the ML-backed quiz.

5.3 The Matching Algorithm

My proposed algorithm based on the results of Chapter 4. The matching algorithm itself could be integrated with any of the classical matching mechanisms that I covered in Chapter 1 to ensure stability, fairness, and efficiency.

The proposed student assignment matching algorithm begins by defining the ministry's five separate rank-order lists of the 43 universities based on their individual score for the algorithmic parameters I, D, Q, C, and U that I defined from their respective, corresponding themes in Study 1: **Infrastructure and Internet**, **Available Departments**, **Quality of Education**, **Campus Environment**, and **University Resources**. In this way, my proposed algorithm models the choice problem for each school individually rather than employing the same

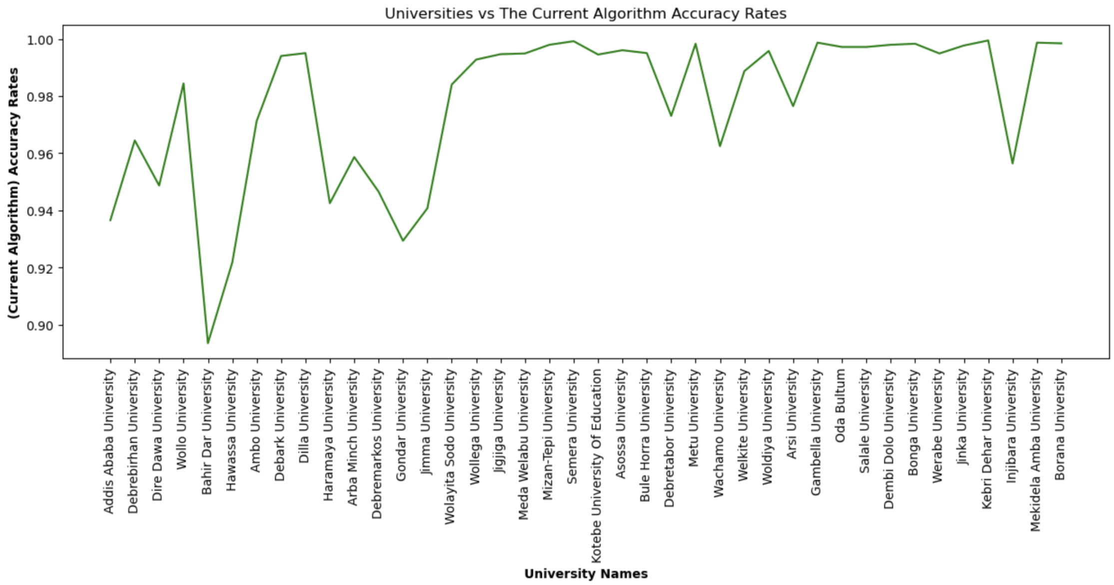


Figure 5.6: Our predicted acceptance rate model achieved a high accuracy rate for each university, ranging from 90% to 99%.

choice rules for many schools [25]. For instance, Jimma University might be #3 in the **Quality of Education**-based rank-order list of the 43 universities, but #10 in the **Infrastructure and Internet**-based rank-order list of the 43 universities.

Then, the algorithm computes the weighted sum rank of each university, based on these characteristic-rankings. The weights would be determined by the appropriate clearinghouse, in this case the MSHE. Let's consider an example; let Jimma University have the following rankings: 20, 10, 3, 2, 1 for the five characteristics (with their respective weights): $I(0.1)$, D (0.2), $Q(0.3)$, $C(0.2)$, and $U(0.2)$. In this case, the weighted sum rank of Jimma Univeristy would be 20(0.1) + 10(0.2) + 3(0.3) + 2(0.2) + 1(0.2) = 5.5, which we would round up to 6. Once the algorithm has computed the weighted sum rank of each university, it generates a sorted list (in ascending order) of the universities based on their weighted sum rank. It breaks ties by prioritizing universities which have a higher **Quality of Education** ranking (this is because that was the characteristic that we identified was most desired by the students in Study 1).

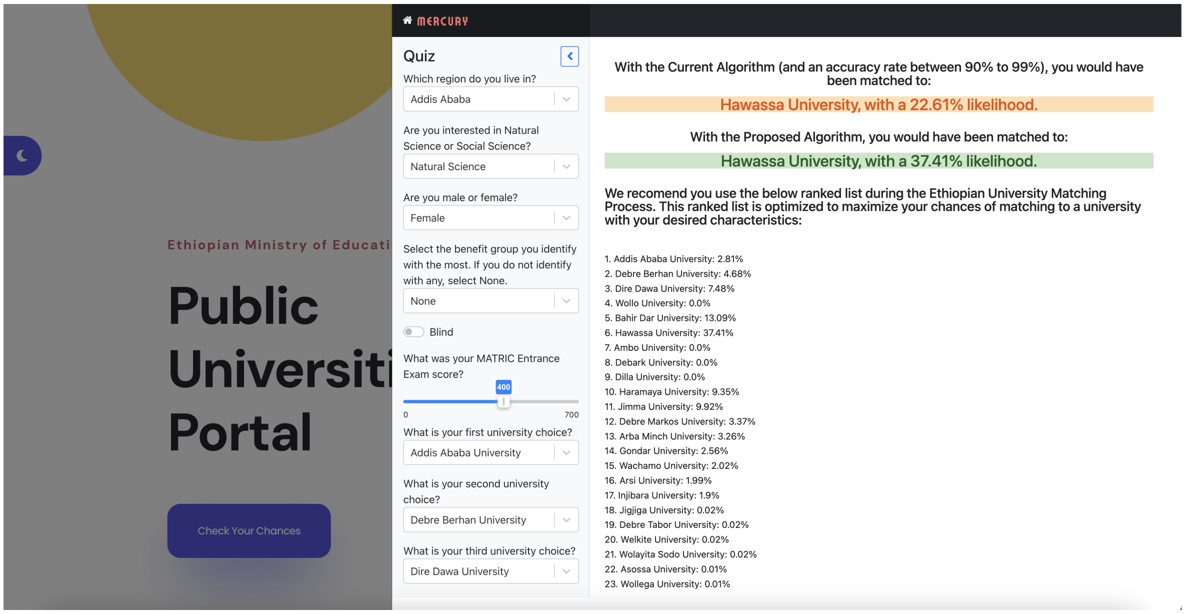


Figure 5.7: The model-generated recommended rank-order list for the given student data.

At this point, the algorithm begins computing the student's acceptance rate for each university based on their top ten universities (which are a proxy for their most desired characteristics), their entrance exam score, and their demographic information. If the student's score and demographics satisfies a hard-to-reach quota requirement for the university, the matching algorithm adds a reward weight to that student's acceptance rate. Then, the algorithm will go through each university in the sorted list of weighted sum ranks of the universities. We want universities with the best characteristics to get their first pick of students. For each university that the algorithm goes through in the sorted list, the algorithm will sort the students in descending order based on their acceptance rates to that university. We also want students with the highest acceptance rates to that university to be chosen first. The algorithm iteratively admit students to that university until all of the available spots for each demographic group of that university have been filled. In other words, we have prioritized students who fulfill hard-to-reach quotas in this process by increasing their acceptance rate (with the reward weight that we mentioned in the prior paragraph). After the algorithm has filled all of the available spots for each demographic group at this

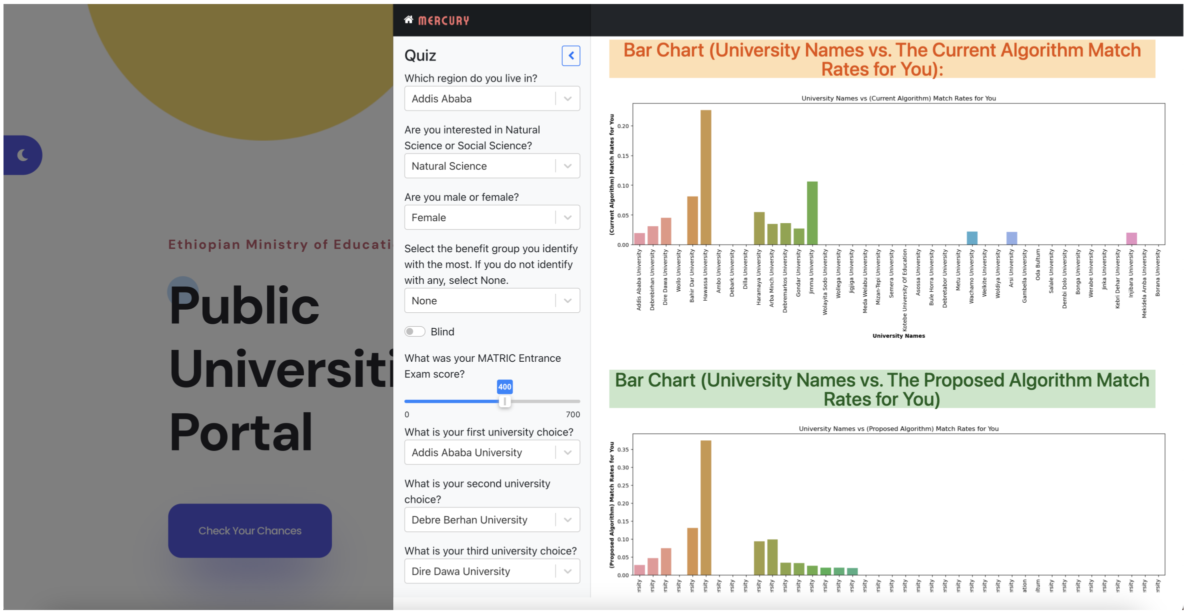


Figure 5.8: The predicted acceptance rates to each university for the current student based on the current algorithm and the proposed algorithm.

university, it will repeat this process with the next university in the sorted list of weighted sum ranks. This algorithm terminates after all the universities have been checked.

The final output of the algorithm will be the matching of admitted students to each university for the 43 universities. While this algorithm does ensure an increased level of fairness in student assignment algorithmic mechanisms, it can also be integrated with any of the existing student placement mechanisms (TTC, Deferred Acceptance, etc.).

Algorithm 1 Multi-Criteria Matching Algorithm

Require: L_s: The student's rank-order list of universities; C_u The most desired characteristics by students (identified in Study 2); $L(u)$: The clearinghouse's rank-order lists of universities per characteristic u; $R(u)$: Reward function for each university c.

Ensure: Stable matching between students and universities that maximizes the students' chances of being assigned to a university with their most desired characteristics, while satisfying the universities' quotas and constraints.

$C(u_Y) \leftarrow$ empty list of weighted characteristic rank of the universities.
for each each university u_Y **do**
  $C(u_Y) \leftarrow \sum_{c \in C_u} w_c \cdot r(c, u_X)$, where w_c is the weight assigned to characteristic c, and $r(c, u_X)$ is the rank of university u_X for characteristic c in $L(u_X)$.
  $C(u_Y).append(u_Y)$.
end for
$SC(u_Y) \leftarrow \text{sort}(\{u_1, \ldots, u_N\}, \text{key} = \lambda u_Y : C(u_Y), \text{reverse=True})$.
for each student s_X **do**
  $G(s_X), E(s_X), D(s_X) \leftarrow$ the given top ten universities (which are a proxy for their most desired characteristics after they have either viewed the website or the PDF), entrance exam score, and quota demographics for student s_X.
  $W_R(s_X) \leftarrow$ reward weights for hard-to-reach quotas of student s_X.
end for
for each student s_X and university u_Y in $L(s)$ **do**
  $A(u_Y, s_X) \leftarrow f(G(s_X, E(s_X), D(s_X), u_Y)$, where f maps the student's top ten universities, entrance exam score, and demographics to an acceptance rate for that university.
  if $len(W_R(s_X)) \geq 1$ **then**
    $A(u_Y, s_X) = A(u_Y, s_X)' \leftarrow$ acceptance rate adjusted upwards to satisfy desired hard-to-reach quotas.
  end if
end for
for each university u_Y in $SC(u_Y)$ **do**
  $T(u_Y) \leftarrow \text{sort}(\{s_1, \ldots, s_N\}, \text{key} = \lambda s_X : A(u_Y, s_X), \text{reverse=True})$, where N is the number of students applying to university u_Y.
  $Q(u_Y) \leftarrow$ number of available spots for each demographic group.
  $A(u_Y) \leftarrow$ empty list of admitted students for university u_Y.
  for each student s_X in $T(u_Y)$ **do**
    if $A(u_Y)$ does not exceed the quota for any demographic group **then**
      $d \leftarrow$ demographic group of student s_X.
      if $Q(u_Y)[d] > 0$ **then**
        $A(u_Y).append(s_X)$.
        $Q(u_Y)[d] \leftarrow Q(u_Y)[d] - 1$.
      end if
    end if
  end for
end for
return the matching of students and universities in $A(u_Y)$.

Chapter 6

Study 2: System Evaluation in Needs-Based Setting

In this study, I built and evaluated the tools that we covered in Chapter 5. These tools were meant to meet the user needs and desires that I identified in my first study, in Chapter 4. This study validates my approach and findings from the first study.

RQ2: How does exposing the student to their personalized acceptance rates to the 43 public universities as well as a recommended rank-order list of the 43 public universities affect the fairness of the matching algorithm?

6.1 Methods

In order to understand the impact of my novel assessment of fairness for student assignment algorithmic mechanisms, I conducted another IRB-approved within-subjects needs-based research study using a randomized controlled trial. I recruited n=40 students from various secondary schools in Ethiopia as my sample. In this study, I deployed a portal intervention that attempted to bridge the gap between algorithm developers and student needs. While using this portal during the intervention condition, students were exposed to comprehensive information about the universities, split up into categories that were directly informed from the results of student needs and values from Study 1, personalized acceptance rates to each university, and an AI-generated recommended rank-order lists of 43 public universities that attempted to help

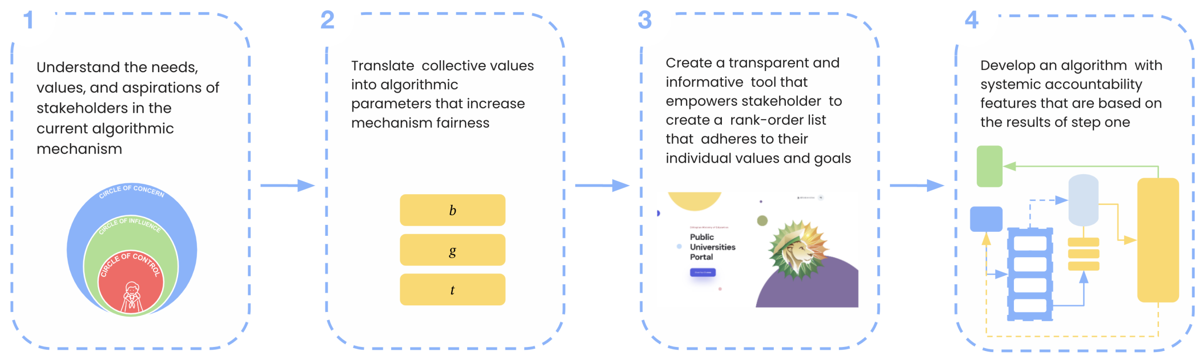


Figure 6.1: Ethos is a system for algorithmic decision making that was designed using a participatory algorithmic design framework. This framework translates the collective values of stakeholders into the parameters of the algorithm. I present this framework as a way to increase fairness, stakeholder empowerment, and systemic accountability in algorithmic decision-making.

students fulfill the requirements of the rank-order list based preference language in student assignment algorithmic mechanisms [59]. I conducted quantitative data analysis in order to measure the performance of the proposed algorithmic mechanism and compare the student's satisfaction of their matching under our proposed algorithmic system versus their predicted matching under the current algorithmic system.

In order to understand the impact of providing high school students with information about the universities, along with personalized acceptance rates and an AI-generated recommended rank-order lists of 43 public universities on the fairness of the centralized university matching process in Ethiopia, this project follows the value-sensitive design methodology by considering the values that are important to the stakeholders involved in the university ranking process [22]. I investigate how the level of information on public universities, including students personalized acceptance rates to each one, affects their rankings, the study aims to uncover how to improve the university ranking process in a way that aligns with students values, needs, and aspirations, as well as those of the Ministry of Science and Higher Education in Ethiopia (MSHE). The use of machine learning to provide students with access to their personalized acceptance rates serves to give them a more personalized and user-centered approach to the university ranking process, aligning with the value of transparency. Through the discussion of the design implications and future work, I will continue to refer back to the values of the stakeholders. My overarching goal in this work is redesigning a student assignment algorithmic mechanism that can be used to improve the university ranking process in a way that is ethical and responsible.

Data Collection

For the control condition, I provided a list of 43 universities in Ethiopia to the students and asked them to submit their initial rank-order list of the universities based on their preferences. Then, I distributed a survey to assess their ranking satisfaction using a 7-point Likert scale. Finally, I ran the Ethiopian Ministry of Science and Higher Education's (MSHE) current student-university assignment algorithm, asked them to submit their finalized rank-order list of the universities, and distributed another survey to assess their matched university satisfaction and trust in the MSHE's current algorithmic mechanism, again using a 7-point Likert scale.

In the intervention condition, I exposed the students to a portal that provides them with comprehensive information about each university that they will be asked to rank. Then, I allowed students to rearrange the rank-order list according to their updated preferences as they went through the website and viewed in-depth details about each of the 43 universities. I also asked each of the students to click a button that would generate an AI-generated recommended rank-order list of universities. Upon clicking this button, the website also displayed where they would have been matched according to the two different student-university assignment algorithms (the MSHE's currently used one and my proposed one). At the end of the intervention, I asked the students to submit their finalized rank-order list of the universities and distributed a survey to assess the students' satisfaction of their matched university, their AI-generated recommended rank-order list, and their trust in my proposed algorithmic mechanism, using 7-point Likert scales.

Data Analysis

For the data analysis, I performed a power analysis over a pilot study of 12 participants in order to determine my sample size. I first analyzed the data using descriptive statistics, including means and standard deviations. Then, I employed Kendall's Tau Distance, as a nonparametric measure of correlation between the rankings of the control and intervention groups, and a paired samples t-test, as an inferential static, to identify significant differences between the means of the two groups.

6.2 Results

Qualitative Analysis

The results show that over 60% of students found the intervention condition to have an impact on their rankings. In this section, I explain my evaluation of this system from a quantitative analysis.

Descriptive statistics of the results for the feedback surveys revealed that the mean trust level and satisfaction level for their likely matched university of the students who used the website intervention were significantly higher than that of students who did not use the website. Students had a 28.3% increase in trust level between the current student assignment algorithmic mechanism and our proposed student assignment algorithmic mechanism. During the control condition, 58.9% of the students stated that they trusted the current student assignment algorithmic mechanism to have their best interests. However, during the intervention condition, 87.2% of the same students stated that they trusted the proposed student assignment algorithmic mechanism to have their best interests.

Inferential statistics through a paired samples t-test indicated that the Kendall's Tau difference in rank-order lists was greater for students after they went through intervention condition ($Mdn = 10.45$, $SD = 17.26$) than after they went through the control condition ($Mdn = 0$, $SD = 0$), $t(12) = 13$, $p = 0.05$, $d = 0.8563$. This led to the results of my power analysis to be interpreted as sufficiently large in order for this study's results to be representative of the millions of students across the nation.

6.3 Discussion

My study suggests that providing students with more information about public universities significantly and positively impacts their rankings and satisfaction with their university matches and their satisfaction with the university selection process. The website intervention tested in the study could be further developed to provide more personalized and informative data to students to further enhance their satisfaction on the 7-Point Likert scales.

Future Work

Future studies could investigate the impact of additional variables, such as report cards on students' university selection processes. Additionally, further research could explore the relationship between the data visualization on the website and the students trust of each of the matching algorithms.

Chapter 7

Future Study: System Evaluation in Assets-Based Setting

The goal of this proposed research study is an assets-based approach that employs a participatory framework to investigate the impact of providing an accessible information service for Ethiopian high school students on the fairness of their assigned university matching. Due to time limitations, I was not able to complete this study. However, I encourage future researchers to take the proposed study and implement it according to their needs.

In particular, in this study, I want to focus on the design process for a future research study that navigates how we can support users in navigating the challenges with the Ethiopian university matching process, even if they do not have their own mobile device or a laptop to run Ethos. To attain the goal of supporting users in navigating their challenges, most researchers prioritize addressing needs through externally-managed ICTs [69]. These processes create ICT solutions that downplay users' agency in devising and pursuing transformational pathways and promote users' dependency on others [69]. In response, researchers are increasingly exploring an assets-based approach to design. At its core, an assets-based approach centers the design process on identifying individuals' and communities' strengths and capacities and exploring feasible ways for users to build on these assets to attain desirable change [69].

Hence, by focusing on the strengths and already-available resources of the students, this proposed study aims to provide them with the necessary information and tools to optimize their outcomes in the university matching process. Namely, we identified that many students had access to Telegram through a prior study in my thesis. This approach empowers the students and recognizes their agency in making informed decisions based on their preferences, values, and aspirations using tools they already have. As we learn from Pei and Nardi, “the asset utilization axis indicates that the more an intervention leverages resources already existing in a population, the more likely the intervention is to have a sustainable impact” [57]. I also want to note the relationship between needs and aspirations: “(the) current lack, which shapes the ‘needs’ of the moment, also impacts the ‘aspirations’ for the future” [47]. In that sense, this study is also supported by Kumar et al.’s definition of aspirations-based design, one property of which is that “aspirations [are] embedded in an interwoven web of pre-existing power structures that was not easily disentangled. . . in this web, aspirations sometimes aligned and, at other times, were in conflict” [47]. Furthermore, the study addresses accessibility barriers, such as language and disability, to ensure that all students can access the information and services provided by the bot intervention.

7.1 Methods

Based on the results of Chapter 4 from my thesis, I propose that this information (personalized acceptance rates and recommended rank-order lists) would optimize the student’s rank-order list because it would lead to an absence of justified envy, lack of information, and misalignment of values. The bot can have one use case, which is outlined in the user flow diagram (see Figure 7.1). It is also parallel to the study in Chapter 6 of my thesis and employs many of the same intervention strategies, variables, and assessments of satisfaction as that study.

For the pre-assessment, the researcher can begin by giving a list of 43 universities in Ethiopia to the students and asked them to send a rank-order list of the 43 public universities based on their preferences, along with their entrance exam score and their demographics. Then, the researcher can run the current centralized university matching process based on the provided information and presented the university matching to the user. Finally, the researcher can distribute a survey to assess their university match satisfaction using a 7-point Likert scale.

Once the pre-assessment is over, the researcher can begin the intervention. First, the researcher can prompt the user to send their needs, aspirations, and desires (from a given list based on subcategories on the aforementioned study in my thesis, such as 'student resources for extracurriculars' or 'desired department'). Then, the bot would send the user a recommended rank-order list of the 43 public universities to optimize their needs, aspirations, and desires. Afterward, the bot can prompt the user to send ten of the 43 provided letter-number combinations for the 43 Public Universities in Ethiopia that they are most interested in matching with. At this point, the researcher can allow the user to rearrange the rank-order list according to their updated preferences as they review the ten PDFs and view in-depth details about those ten universities. The ten PDFs would also include a section with the user's personalized acceptance rate for each university. Then, the researcher can prompt the user to re-send their ROL. After this step, the bot can run my proposed matching algorithm and presented the new university matching to the user. Finally, the researcher's bot can prompt users to send their satisfaction with the new university matching on a 7-point Likert scale.

To develop the Telegram bot, the researcher can write a Python script and integrate the Telegram Bot API. For now, I have included a high-fidelity Figma prototype of the potential bot (see Figure 7.2). A critical aspect of the bot's design was to ensure that it would be an accessible interface for as many users as possible. More specifically, I concentrated on how to receive accurate inputs from users that can be difficult to reach, such as those in vulnerable communities, with limited resources, or who have accessibility difficulties. A couple of specific examples of these vulnerable communities are those with language barriers or familial obligations which keep them close to home. My service took a couple of different steps to address language barriers and accessibility difficulties. To address language barriers, the bot provides an option to translate the text in the PDFs and the bot's messages from English into any one of the three most commonly used languages out of the 85 languages in Ethiopia: Amharic, Oromo, and Tigrinya. To address accessibility barriers, for those with typing difficulties or visual impairments, the bot supports voice input and output options through Telegram's TalkBack feature on Android and VoiceOver feature on iOS. Finally, to address familial obligations which keep the user close to home, the bot can ask the user where they currently live and integrates this information into the recommended rank-order list which it provides the user. Additionally, to ensure users get the information they need and integrate iteration into the bot's development cycle, the bot will include elements like error handling and user feedback.

In order to send the PDFs of the universities to the user in the third step of the intervention, the researcher can compiled a database of PDFs for every university and mapped each one to a specific letter-number combination. For instance, if the user inputs the code “A1” as one of their ten universities which they would like a PDF for during the intervention part of the study, they will receive a compiled PDF with detailed and factual information on that university (as well as the other nine universities) that also includes their predicted likelihood of acceptance.

To summarize, this proposed research study tests the effects of an accessible Telegram bot service that aims to provide comprehensive information on the 43 public universities in Ethiopia. This service can help high school students through the centralized university matching process by providing them with the information that they need on the universities to provide an accurate and truthful rank-order list of the 43 public universities. The purpose of this study is to ensure an accessible information service that aligns with the local community’s resource access to help users navigate their own challenges with the Ethiopian centralized university matching process. In other contexts, this might mean developing low-technology solutions or partnering with the local community to provide a physical location to gather user feedback on the proposed intervention. Society-impacting algorithms that operate in isolation from the community they affect are in danger of harming them. They must operate in symphony with the society which they impact. Therefore, my goal with proposing this assets-based study was to give agency back to the community during the university matching process using the tools they already have on hand.

7.2 Discussion

Future work includes expanding the scope of the bot beyond the 43 public universities in Ethiopia and integrating it with other education-related services to provide a comprehensive educational resource for high school students in Ethiopia. More specifically, this bot can integrate study resources and advice for the students on the entrance exam, which will help their chances of receiving their most optimal university placement. In addition, the bot can be expanded to support the other 82 languages spoken in Ethiopia to increase accessibility for all students.

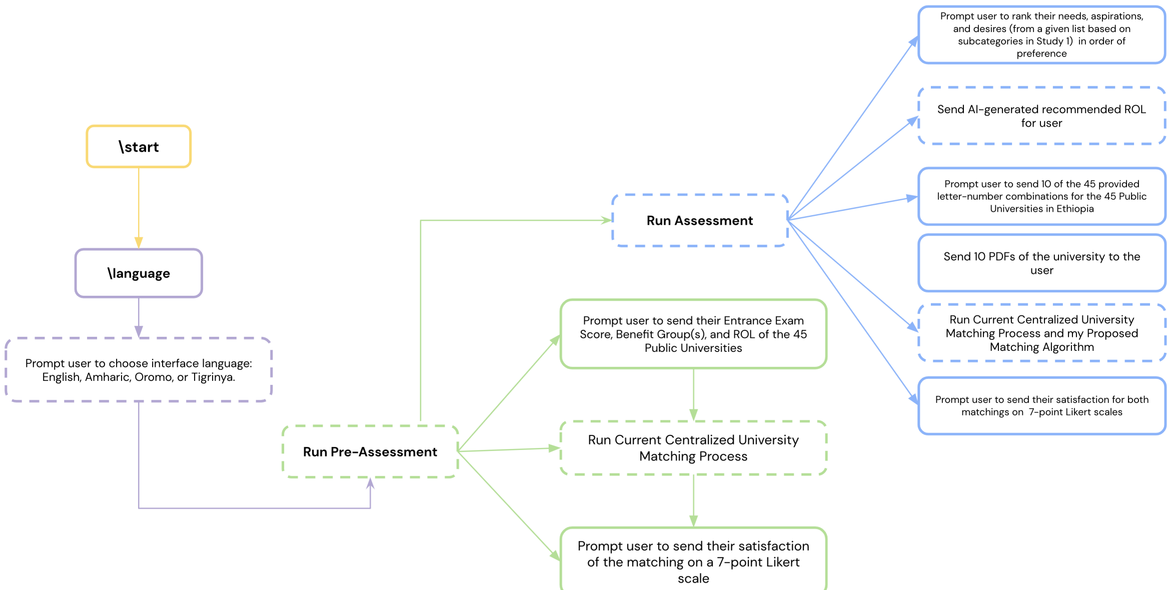


Figure 7.1: The user flow detailing how end users can interact with the bot service.

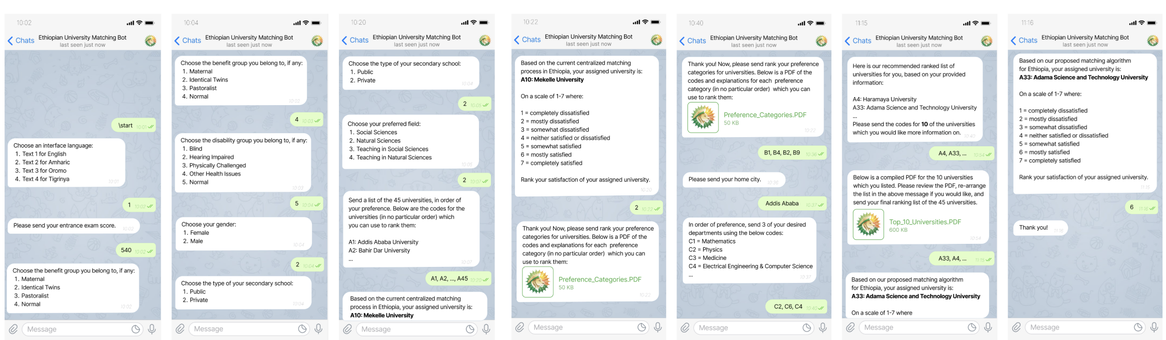


Figure 7.2: The high-fidelity wireframe of the bot service.

Chapter 8

Discussion

In this thesis, I took the case of student university assignments in Ethiopia as a case study to implement an end-to-end participatory algorithm design process. The result was Ethos, my proposed student assignment algorithmic mechanism, which uses deliberative democracy [10, 5] to translate the collective needs and values of stakeholders into algorithm parameters that increase mechanism fairness and empower stakeholders. Ethos also holds each of the universities in the system accountable by customizing the choice problem for each school rather than applying the same choice rules for many schools [24].

Challenging the ecosystem the algorithm will live in is essential and one of the core findings of this thesis. In the case of this research, this meant questioning and iterating on the preference elicitation process during data collection. I had to slow down the algorithm design process in order to stop and listen to the participants. This led to a bottom-up approach during my algorithm design process which included traveling overseas in Study 1 in order to conduct the ethnographic research. This slowed down approach to participatory algorithm design was pivotal in helping us to identify the needs and desires of the agents controlling the algorithm output: the students.

In Study 2, I found that the level of information available on public universities had a significant effect on students rankings. Specifically, students who used the website intervention showed higher satisfaction levels with their university matches and trust in the centralized university process than those who did not use the website. An advantage of this study was that I was able to assess several different dimensions of student assignment and matching under a controlled setting.

8.1 Recommendations

Design stakeholder input into the algorithm design process.

A core value in this work is showing how, as algorithmic mechanism design researchers, we must resist techno-determinism [36] and increase stakeholder empowerment by integrating deliberative democracy [11] and value-sensitive design [22] into our algorithm design process. This requires awareness of participatory mechanisms and developing methods to expand participatory methods to incorporate complex algorithm design challenges. A central aspect of this work will be to allow the people affected by the system to audit and modify the algorithm by designing human-centered feedback and equity into the mechanism from the beginning rather than trying to patch up a broken system with an algorithm. I followed this principle in Study 1, when I took in student feedback on values to consider for the resulting novel value-centered matching algorithm that I tested with students in Study 2.

Build in community.

Leading with grounded theory in my research meant constantly and intentionally questioning the ecology that the algorithm would exist in and being in community with the participants in order to incorporate their participation in the proposed algorithm's research and design. This was prior to any data analysis or software development. The extra time and deliberate intentionally I took to understand the user perspectives early on in my research, during Study 1, was necessary and led my research to better suit them, which I witnessed in Study 2. My goal with this thesis was to create an algorithm which encourages feedback loops, redistributes the power imbalance between the algorithm and the agent, and ultimately, gives agency back to the community.

8.2 Limitations

One limitation of Study 1 is that I only recruited participants from one geographic area in Ethiopia (Addis Ababa). Despite the diverse set of socioeconomic levels which I tried to recruit from, there is still a stark and disproportionate amount of privilege that students in the capital city have compared to the lesser developed and rural regions of Ethiopia. A limitation of

Study 2 is that it had low ecological validity because the controlled conditions in this experiment are not necessarily what happen in the real world.

8.3 Positionality and Reflexivity

I cannot end this thesis without expressing how much I have truly enjoyed this research process. As an Ethiopian-born researcher, this research represented much more than an intellectual pursuit. Researching and building an equitable algorithmic system, Ethos, based on the lived experiences of people from my homeland carries deep and symbolic meaning. I feel a deep sense of gratitude to all of the students who participated in Study 1 and Study 2. Without their honest contributions, Ethos would not have been possible. This research advocates for amplifying stakeholders' voices in algorithmic design, urging researchers to reflect on their role in perpetuating broken systems and prioritizing societal impact by focusing on defending users from flawed social systems. This project exemplifies an effort to amplify the voices of participants to lead the direction of the algorithmic research. I also have a tremendous amount of hope and respect for the MSHE in considering the infrastructural challenges this liberation would require. Finally, I am excited for the future of the system that they build for the secondary education to higher education pipeline in Ethiopia.

Chapter 9

Conclusion

From Study 1, I present the following two design implications for future research in student assignment participatory algorithm design: (1) in order for participants to be able to make informed rankings, they need accessible information, and (2) in order for universities to get diverse student bodies, they need to explain to the students how their university adheres to the student's needs, aspirations, and desires. Following the results of Study 2, we learn that students consider a portal and predicted acceptance rate model as necessary tools for any student assignment mechanism because they empowered them when making their rank-order lists of universities. In this study, students described how these "transformative" tools provided them with crucial information about their chances of acceptance which would help to comfort them wherever they matched, instill their confidence in the matching process, and ultimately contribute to their happiness and success.

My thesis is a real-world research application of enacting theories of justice in the algorithm design process by gathering the needs, aspirations, and desires of the people affected by the algorithm. My research utilizes participatory design approaches for the development of an algorithmic system for school assignment in Ethiopia. Through ethnographic research with the people who are subject to decisions by the system, I developed a theory of fairness in this system. I used these insights to develop and evaluate a technical system based on the desires and values of the participants.

I also want this project to exemplify an effort to amplify the voices of users in leading the direction of algorithmic research. Imagine if we continuously paused, grounded, and re-validated our research by asking ourselves: in the best case scenario where this research benefits the subject, what is our role? are we perpetuating a broken system with a band-aid solution under the guise of algorithmic novelty? We are researching solutions which could be the last line of defense between the user and a broken social system, so let us be encouraged to be conscious of our sociotechnical imaginary as we research algorithmic solutions.

Bibliography

[1] Atila Abdulkadiroglu et al. *Minimizing justified envy in school choice: the design of New Orleans' OneApp*. Tech. rep. National Bureau of Economic Research, 2017.

[2] Atila Abdulkadiroğlu and Tayfun Sönmez. "School choice: A mechanism design approach". In: *American economic review* 93.3 (2003), pp. 729–747.

[3] Atila Abdulkadiroğlu et al. "Efficiency, justified envy, and incentives in priority-based matching". In: *American Economic Review: Insights* 2.4 (2020), pp. 425–442.

[4] Nikhil Agarwal. "An empirical model of the medical match". In: *American Economic Review* 105.7 (2015), pp. 1939–1978.

[5] Ari-Veikko Anttiroiko. "Building strong e-democracy: the role of technology in developing democracy for the information age". In: *Communications of the ACM* 46.9 (2003), pp. 121–128.

[6] Kenneth J Arrow. *Social choice and individual values*. Vol. 12. Yale university press, 2012.

[7] Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. *Handbook of social choice and welfare*. Vol. 2. Elsevier, 2010.

[8] Michel Balinski and Tayfun Sönmez. "A tale of two mechanisms: student placement". In: *Journal of Economic theory* 84.1 (1999), pp. 73–94.

[9] Jack Bandy. "Problematic machine behavior: A systematic literature review of algorithm audits". In: *Proceedings of the acm on human-computer interaction* 5.CSCW1 (2021), pp. 1–34.

[10] Benjamin Barber. *Strong democracy: Participatory politics for a new age*. Univ of California Press, 2003.

[11] Derek WM Barker. "Deliberative justice and collective identity: A virtues-centered perspective". In: *Political Theory* 45.1 (2017), pp. 116–136.

[12] Anna Maria Barry-Jester, Ben Casselman, and Dana Goldstein. “The New Science of Sentencing”. In: *The Marshall Project* (2015). URL: `https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing`.

[13] Ruha Benjamin. “Assessing risk, automating racism”. In: *Science* 366.6464 (2019), pp. 421–422.

[14] Ruha Benjamin. *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life*. Duke University Press, 2019.

[15] Ruha Benjamin. “Quantified Racism”. In: *The Quantified Worker: Law and Technology in the Modern Workplace* (2023), p. 293.

[16] Anna Brown et al. “Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12.

[17] Jenna Burrell. “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”. In: *Big data & society* 3.1 (2016), p. 2053951715622512.

[18] Yves Cabannes. “Participatory budgeting: a significant contribution to participatory democracy”. In: *Environment and urbanization* 16.1 (2004), pp. 27–46.

[19] Robert Chambers. “Poverty and livelihoods: whose reality counts?” In: *Environment and urbanization* 7.1 (1995), pp. 173–204.

[20] Yan Chen and Tayfun Sönmez. “School choice: an experimental study”. In: *Journal of Economic theory* 127.1 (2006), pp. 202–231.

[21] Hao-Fei Cheng et al. “Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders”. In: *Proceedings of the 2019 chi conference on human factors in computing systems*. 2019, pp. 1–12.

[22] Sasha Costanza-Chock. *Design justice: Community-led practices to build the worlds we need*. The MIT Press, 2020.

[23] Fernando Delgado et al. “Stakeholder Participation in AI: Beyond" Add Diverse Stakeholders and Stir"”. In: *arXiv preprint arXiv:2111.01122* (2021).

[24] Battal Doğan. “Mechanism Design Approach to School Choice: One Versus Many”. In: *The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers*. Springer, 2019, pp. 469–472.

[25] Federico Echenique and M Bumin Yenmez. "How to control controlled school choice". In: *American Economic Review* 105.8 (2015), pp. 2679–2694.

[26] Haluk Ergin and Tayfun Sönmez. "Games of school choice among boston students". In: *Manuscript. MIT and Koç University* (2003).

[27] Florian Evequoz et al. "Diverse Representation via Computational Participatory Elections-Lessons from a Case Study". In: *Equity and Access in Algorithms, Mechanisms, and Optimization*. 2022, pp. 1–11.

[28] Gregory Ewing and Ibrahim Demir. "An ethical decision-making framework with serious gaming: a smart water case study on flooding". In: *Journal of Hydroinformatics* 23.3 (2021), pp. 466–482.

[29] Martin Fisher. "Income is development: KickStart's pumps help Kenyan farmers transition to a cash economy". In: *Innovations: Technology, Governance, Globalization* 1.1 (2006), pp. 9–30.

[30] Rachel Freedman et al. "Adapting a kidney exchange algorithm to align with human values". In: *Artificial Intelligence* 283 (2020), p. 103261.

[31] David Gale and Lloyd S Shapley. "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1 (1962), pp. 9–15.

[32] Marzyeh Ghassemi et al. "A review of challenges and opportunities in machine learning for health". In: *AMIA Summits on Translational Science Proceedings* 2020 (2020), p. 191.

[33] Steven Glazerman and Robert H Meyer. "Public school choice in Minneapolis, midwest approaches to school reform". In: *Proceedings of a Conference Held at the Federal Reserve Bank of Chicago*. 1994, pp. 26–27.

[34] Rustamdjan Hakimov and Onur Kesten. "The equitable top trading cycles mechanism for school choice". In: *International Economic Review* 59.4 (2018), pp. 2219–2258.

[35] Matt Haney, Stevon Cook, and Rachel Norton. "Developing a Community Based Student Assignment System for SFUSD". In: *San Francisco Unified School District. Retrieved April* 13.2020 (2018), pp. 2019–20.

[36] Robert L Heilbroner. "Do machines make history?" In: *Technology and culture* 8.3 (1967), pp. 335–345.

[37] Yoan Hermstrüwer. "Transparency and fairness in school choice mechanisms". In: *MPI Collective Goods Discussion Paper* 2019/11 (2019).

[38] Zoë Hitzig. "The normative gap: mechanism design and ideal theories of justice". In: *Economics & Philosophy* 36.3 (2020), pp. 407–434.

[39] Melissa R Ho, Emmanuel K Owusu, and Paul M Aoki. "Claim Mobile: Engaging conflicting stakeholder requirements in healthcare in Uganda". In: *2009 International conference on information and communication technologies and development (ICTD)*. IEEE. 2009, pp. 35–45.

[40] Melissa R Ho et al. "Human-computer interaction for development: The past, present, and future". In: *Information Technologies & International Development* 5.4 (2009), pp–1.

[41] Caroline M Johnston, Simon Blessenohl, and Phebe Vayanos. "Preference elicitation and aggregation to aid with patient triage during the covid-19 pandemic". In: *Workshop on Participatory Approaches to Machine Learning*. 2020.

[42] Anson Kahng et al. "Statistical foundations of virtual democracy". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 3173–3182.

[43] Onur Kesten. "On two kinds of manipulation for school choice problems". In: *Economic Theory* 51 (2012), pp. 677–693.

[44] Onur Kesten. "Student placement to public schools in the US: Two new solutions". In: *Unpublished mimeo, University of Rochester* 9 (2004).

[45] Fuhito Kojima, M Utku Ünver, et al. "The" Boston" School-Choice Mechanism". In: (2010).

[46] Amanda Kube et al. "Just resource allocation? How algorithmic predictions and human notions of justice interact". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. 2022, pp. 1184–1242.

[47] Neha Kumar et al. "Aspirations-based design". In: *Proceedings of the tenth international conference on information and communication technologies and development*. 2019, pp. 1–11.

[48] Heidi Ledford. "Millions affected by racial bias in health-care algorithm". In: *Nature* 574.31 (2019), p. 2.

[49] Min Kyung Lee et al. "Participatory algorithmic management: Elicitation methods for worker well-being models". In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021, pp. 715–726.

[50] Min Kyung Lee et al. "WeBuildAI: Participatory framework for algorithmic governance". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–35.

[51] Lydia T Liu et al. "Lost in translation: Reimagining the machine learning life cycle in education". In: *arXiv preprint arXiv:2209.03929* (2022).

[52] Cecilia Machado and Christiane Szerman. "Centralized admission and the student-college match". In: (2016).

[53] Thayer Morrill. "Two simple variations of top trading cycles". In: *Economic Theory* 60.1 (2015), pp. 123–140.

[54] Ritesh Noothigattu et al. "A voting-based system for ethical decision making". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018.

[55] Parag A Pathak. "The mechanism design approach to student assignment". In: *Annu. Rev. Econ.* 3.1 (2011), pp. 513–536.

[56] Parag A Pathak. "What really matters in designing school choice mechanisms". In: *Advances in Economics and Econometrics* 1.12 (2017), pp. 176–214.

[57] Lucy Pei and Bonnie Nardi. "We did it right, but it was still wrong: Toward assets-based design". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–11.

[58] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. "Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–14.

[59] Samantha Robertson et al. "Expressiveness, Cost, and Collectivism: How the Design of Preference Languages Shapes Participation in Algorithmic Decision-Making". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–16.

[60] Alvin E Roth and Elliott Peranson. "The redesign of the matching market for American physicians: Some engineering aspects of economic design". In: *American economic review* 89.4 (1999), pp. 748–780.

[61] Devansh Saxena et al. "A human-centered review of algorithms used within the US child welfare system". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–15.

[62] Lloyd Shapley and Herbert Scarf. "On cores and indivisibility". In: *Journal of mathematical economics* 1.1 (1974), pp. 23–37.

[63] Thomas N Smyth et al. "Where there's a will there's a way: mobile media sharing in urban india". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2010, pp. 753–762.

[64] Tayfun Sönmez and M Utku Ünver. "Matching, allocation, and exchange of discrete resources". In: *Handbook of social Economics*. Vol. 1. Elsevier, 2011, pp. 781–852.

[65] Sharifa Sultana et al. "Design within a patriarchal society: Opportunities and challenges in designing for rural women in bangladesh". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[66] Halil Toros and Daniel Flaming. "Prioritizing homeless assistance using predictive algorithms: an evidence-based approach". In: *Cityscape* 20.1 (2018), pp. 117–146.

[67] Kentaro Toyama. *Geek heresy: Rescuing social change from the cult of technology*. PublicAffairs, 2015.

[68] Michael Veale, Max Van Kleek, and Reuben Binns. "Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making". In: *Proceedings of the 2018 chi conference on human factors in computing systems*. 2018, pp. 1–14.

[69] Marisol Wong-Villacres et al. "Reflections on assets-based design: A journey towards a collective of assets-based thinkers". In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW2 (2021), pp. 1–32.

[70] Mohammad Yaghini, Andreas Krause, and Hoda Heidari. "A human-in-the-loop framework to construct context-aware mathematical notions of outcome fairness". In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021, pp. 1023–1033.

[71] Takuya Yokota and Yuri Nakao. "Toward a decision process of the best machine learning model for multi-stakeholders: a crowdsourcing survey method". In: *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 2022, pp. 245–254.

[72] Angie Zhang et al. "Deliberating with AI: Improving Decision-Making for the Future through Participatory AI Design and Stakeholder Deliberation". In: *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW1 (2023), pp. 1–32.

[73] Angie Zhang et al. "Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–19.

[74] Douglas Zytko et al. "Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains". In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 2022, pp. 1–4.

Ethos: Designing Algorithmic Mechanisms for Increased Fairness, Stakeholder Empowerment, and Systemic Accountability

LIYA MULUGETA, University of California, Berkeley, USA
NILOUFAR SALEHI, University of California, Berkeley, USA

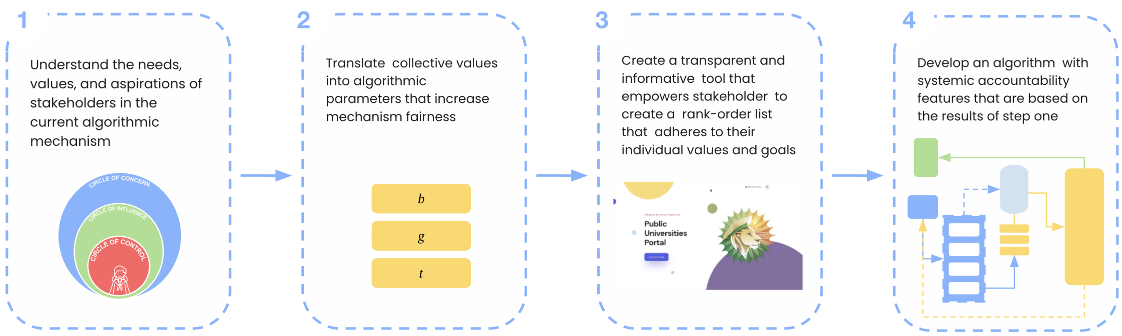


Fig. 1. Ethos is a system for algorithmic decision making that was designed using a participatory algorithmic design framework. Our system translates the collective values of stakeholders into the parameters of the algorithm. We present this framework as a way to increase fairness, stakeholder empowerment, and systemic accountability in algorithmic decision-making.

A key concern of policymakers who use student-assignment algorithms is increasing diversity and fairness. Drawing on two years of research into the student assignment algorithmic mechanism of Ethiopia, we study how the collective values of stakeholders can shape the parameters of algorithmic mechanisms. We find that prioritizing the collective values of stakeholders in the initial stages of the algorithm design process enhances university diversity, empowers students to submit truthful rank-order lists, and establishes accountability mechanisms for universities. One of the main contributions in this paper is a novel assessment of mechanism fairness, as defined by having: (1) an absence of **justified envy**, (2) an absence of **lack of information**, and (3) an absence of **misalignment of values**. Ethos, the core technical contribution in this paper, consists of a machine learning-backed acceptance rate quiz for Ethiopian public universities, an informative portal for Ethiopian students, and a generalized student-university matching algorithm. Our process for designing and evaluating Ethos consists of two in-person user studies (Study 1, n=33; Study 2, n=40) in which we identify and assess the real-world impact of our system and algorithm parameters. We argue that listening to and prioritizing the collective values of stakeholders is critical to building diverse and fair algorithmic mechanisms and offer generalizable methods for doing so.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**.

Additional Key Words and Phrases: participatory design, mechanism design, algorithmic mechanisms, fairness, user empowerment, student assignment

Authors' addresses: Liya Mulugeta, liyak@berkeley.edu, University of California, Berkeley , Berkeley, California, USA; Niloufar Salehi, nsalehi@berkeley.edu, University of California, Berkeley , Berkeley, California, USA.

ACM Reference Format:
Liya Mulugeta and Niloufar Salehi. 2023. Ethos: Designing Algorithmic Mechanisms for Increased Fairness, Stakeholder Empowerment, and Systemic Accountability. 1, 1 (November 2023), 19 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Algorithmic mechanisms are increasingly being adopted within decision-making across various public sector domains such as healthcare, child welfare, and law enforcement [10, 14, 29, 55, 59, 61]. These systems impact people's lives and have faced numerous criticisms from researchers over their discriminatory practices [12, 13] and unintended consequences to vulnerable communities [11, 44]. There is a growing concern that systems which are designed without consideration to the needs and values of the affected communities are in danger of harming them [20]. A growing body of literature on participatory algorithm design implements social choice theory as a framework for these systems to collect and aggregate individual stakeholders' preferences [6, 16, 24, 25, 27, 38, 39, 43]. However, the path from gathering and aggregating individual stakeholder preferences to providing an algorithmic mechanism that can be practically implemented is not straight-forward [45, 46, 49, 62, 63]. In this paper, we employ a bottom-up approach to design an algorithmic mechanism whose parameters were shaped by the collective values of stakeholders.

In several countries, including Chile, Turkey, Germany, Taiwan, and the U.K., a high-stakes application of algorithmic mechanisms is the national assignment of students to public universities through a centralized university matching process [2, 47, 58]. In this system, students are assigned to a university among a limited number of universities. A student may also be assigned to zero universities. We researched the student assignment algorithmic mechanism in Ethiopia with the aim of developing a novel algorithmic mechanism which increased the diversity of its public universities' student bodies. This case study offers a practical algorithmic mechanism that is rooted in value-sensitive design and emphasizes user empowerment by directly using the collective values of the stakeholders to set the values of the algorithm which will facilitate their allocation.

Our process for designing and evaluating Ethos consists of two in-person user studies (Study 1 and Study 2) with a total of 73 participants. The purpose of Study 1 was to identify the collective needs, values, and aspirations of students who undergo the national student assignment algorithmic mechanism in Ethiopia. For Study 1, we conducted 33 semi-structured interviews with students across public and private secondary schools in Ethiopia during Winter 2022. The purpose of Study 2 was to evaluate the diversity and fairness outcomes of our proposed algorithmic mechanism. For Study 2, we conducted a within-subjects needs-based study using a randomized controlled trial in Summer 2023. We recruited 40 students from various regions and secondary schools in Ethiopia as our sample. This study deployed an informative portal and our proposed student-university algorithmic mechanism that can be generalized to other contexts and countries. Our combination of qualitative and quantitative methods across both of these studies allowed us to triangulate our research so that we could have a stronger understanding of the student assignment algorithmic mechanism for Ethiopia.

Mechanism fairness determines how the preferences of the students are aggregated and how the matching is performed. Traditionally, mechanism fairness is based on the absence of justified envy [3, 40]. A student has justified envy for another student assigned to a university if they prefer the university to the one that they are assigned to and they have higher priority for it than at least one student who is placed to it [40]. In our work, we define a novel assessment of mechanism fairness as having: (1) an absence of **justified envy**, (2) an absence of **lack of information**, and (3) an absence of **misalignment of values**. The second property arises from Study 1, where we found that most students randomly ranked 90% of the universities on their rank-order list, in large part due to a lack of information on

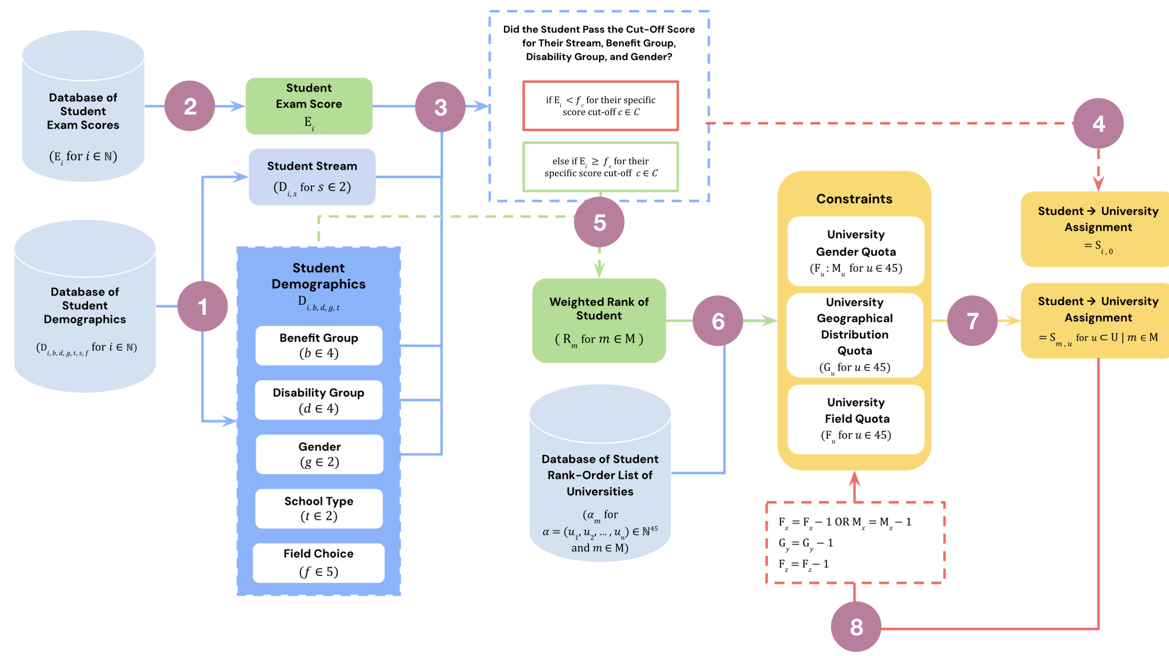


Fig. 2. Model of the Current Centralized University Matching Process in Ethiopia. The possible rank-order lists of the 43 public universities are represented by the n-ordered tuples $\alpha = (u_1, u_2, ..., u_n) \in \mathbb{N}^{43}$ (the set of all 43-ordered tuples of natural numbers). Here is the sequence of events to assign one student from the database of students to a university: (1) find the current student's demographics, (2) find the current student's exam scores, (3) check if the current student passes the score cutoff for their demographics, (4) if the current student's score is below the cutoff, set their student-university assignment to no universities, (5) else if the current student's score is above the cutoff, assigned a weighted rank to the student based on their score and demographics, (6) find the current student's rank-order list of universities and gather the constraints of the universities, (7) assign the current student to their highest-ranking university that also satisfies the university's quota requirements, (8) remove this student from the quota requirement of their assigned university.

them. The third property is also informed from our analysis in Study 1, where we found that the university qualities need to align with the needs, desires, and aspirations of the students.

We analyze the impact of this novel definition of fairness on student assignment algorithmic mechanisms from our results in Study 2. In this study, we deploy our proposed system, Ethos. Ethos consists of three main technical components: 1) an information portal, 2) a machine learning-based system for suggesting universities, 3) a multi-criteria matching algorithm. The machine learning model used in Ethos can predict a given student's acceptance rates to each university based on the Ministry of Science and Higher Education's (MSHE) current student-assignment algorithm. We built this model on over 160,000 data points of previous student-university assignments. To develop it, we used one-hot encoding and max-voting on an ensemble of three different models.

The ethos of this work is that building more equitable societal mechanisms requires critically examining our technology's systemic impacts and its positionality within a community in supporting or suppressing user agency. In this work, we worked to respect this concept by listening to the students on their drawbacks, desires, and aspirations for the student assignment algorithmic mechanism in Study 1 and using these results to inform the production of our algorithmic mechanism in Study 2. We showcase how the normative gap in the current student assignment algorithmic mechanism is that it does not align with the actual demands for justice for more diverse student bodies at the Ethiopian public universities. This is because it neglects to consider the student's Circles of Influences which might deter them from submitting truthful and accurate rank-order lists of the universities. Within the context of mechanism design, which is the backbone of algorithmic mechanisms such as student-university assignment, this normative gap disrupts the classical definition of mechanism fairness. Ultimately, we argue that expanding our sociotechnical imaginary requires avenues that give agency to users in shaping algorithmic systems which affect them.

2 BACKGROUND

In student assignment algorithmic mechanisms, each university has a certain number of available seats. For each university, there is a strict priority order of all the students, and each student has a strict order of preferences for all the universities [28]. The final student-university allocations are allocated by a clearinghouse through a classical student assignment algorithm that is rooted in mechanism design [4, 50, 56]. These clearinghouses differ from country to country, from the Ministry of Education in one country to a private organization in another country [40]. In the U.S., for example, the placement of medical students to residency options is determined by a clearinghouse called the National Residency Match Program [54]. Additionally, many cities in the U.S., such as San Francisco and Boston, also delegate clearinghouses to assign students to primary and secondary schools [2, 42, 53]. In recent years, these classical student assignment algorithms have been researched and modified with the hope that the newer versions will increase diversity within schools and optimize the outcomes of all of the stakeholders [1, 7, 31, 34, 41, 42, 48]. However, in practice, these algorithms have several limitations and instead decreased diversity in schools [32, 53].

The clearinghouse for Ethiopia's national student assignment mechanism is the Ministry of Science and Higher Education in Ethiopia (MSHE), which has a set of criteria to admit and place students in higher institutions, including student entrance exam minimum requirements and university quota constraints. We developed a high-level system model (see Figure 2) to outline the input parameters, constraints, and output (set of student-university assignments) according to a specification document provided by the MSHE[1]. From this high-level system diagram and data, we were able to create a machine learning model which can predict the acceptance rate of a given student to any university in the current centralized matching process with a high-level of accuracy (over 90%).

3 RELATED WORK

3.1 Student Assignment Algorithms

Existing literature in student assignment algorithms builds upon two core matching mechanisms [2, 48, 51], Gale and Shapley's Deffered Acceptance algorithm [28] and Gale's Top Trading Cycles (TTC) [56]. One of the most frequently-used real-life student assignment algorithms is the Boston Student Placement algorithm [2, 42]. This algorithm and its minor modifications are presently implemented in various cities including Boston, Seattle, Minneapolis, Lee County, and Florida [2]. This algorithmic mechanism begins by considering, for each school, all the students who listed that school as their first choice and allocating them by their priority order one at a time until there are no available seats left or there are no students who have listed it as their first choice [40]. Next, consider the subset of students who have not been placed in any school in the previous step. For each school with available seats, only those students who have listed the school as their second choice are considered. The priority order of students within each school is again followed to

[1]The Ministry had legal and privacy protections which limited them from sharing the code behind the current matching algorithm with the first author, but shared other details and data about the process.

place these students one at a time until either no seats are left or there are no students remaining who have listed the school as their second choice. This cycle repeats until all students are placed.

The Boston algorithm has limitations [40]. Specifically, the algorithm is not strategy-proof, as it can give students incentives to misrepresent their true preferences in order to maintain their priority for certain schools. The algorithm's design forces students to think strategically and make non-truthful submissions, which can lead to suboptimal outcomes. In addition, school district authorities often advise students and their parents to make strategic choices [30]. Empirical evidence has confirmed the strategic behavior of students under the Boston algorithm. Chen and Sonmez conducted an experiment and found that 80% of the subjects chose to misrepresent their preferences under the algorithm [18]. This is due to the fact that students fear losing their priority for certain schools and believe that strategic choices will increase their chances of being assigned to their preferred school. Misstating preferences is so prevalent that even suggestions encouraging such behavior have appeared in the press [23]. The high incidence of strategic behavior in the Boston algorithm raises concerns about this mechanism's fairness.

3.2 Theories of Justice

In "Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms," the authors came up with four design implications for student assignment systems, including: providing relevant and accessible information, aligning and realigning algorithmic objectives with community goals in mind, reconsidering how stakeholders express their needs and constraints, and making appropriate,reliable avenues for recourse available [53]. One of the main takeaways of this paper was that student assignment algorithms exist within and to uphold a political ideology that privileges individual choice sometimes at the cost of other values, such as democracy, resource equality, and desegregation.

This takeaway aligns with Hitzig's proposal that mechanism design enacts a theory of justice [35]. Hizig bases this proposal analysis on two unusual features of the Boston algorithm redesign: 1) that the economic theory is enacted in the school system and 2) that it draws on an elaborate, but unarticulated, normative framework. Putting these two features together suggests that mechanism design can be reframed through an ideal theory of distributive justice. From this, she argues that there is a normative gap between the implicit normative theory of the mechanism and the actual demands of justice.

3.3 Participatory Design and User Empowerment

There has been growing interest in incorporating distributive justice and individual preferences into algorithmic decision-making through the use of participatory algorithm design [14, 19, 45, 64–66]. Our research relies heavily on the framing which participatory design provides. Incorporating this framework into our research is necessary for evaluating how students are choosing to participate in this school choice mechanism.

Toyama's "amplification thesis," posits that technology's only impact is amplification [60]. Therefore, a major consideration should be: what is amplified? Decision-making systems that are built and operated without input from the community they affect are in danger of harming them [20]. The concept of technology giving agency back to its users is well-articulated in Chambers' work, where he argues that researchers must respect the basic human right of vulnerable communities to conduct their own analysis and listen to their inputs as we research a solution for that community [17]. In this work, we respect this concept by listening to the students on their drawbacks, desires, and aspirations for the centralized university matching process in Study 1. Additionally, Burrell's Material Eco-Systemic Approach defines an improved ethic of design in ICT where ICT methods should account for, support, and amplify the agency of its users [15]. By extension, we can deduce that the true value of technology in resource-constrained settings is how much it can empower and liberate its users, such as in the case study of Fisher's KickStart [26]. In order to achieve this end goal, our work employs a four-step research method: (1) Ethnographic Research, (2) Need-Finding Research, (3) Iterate and Refine Prototypes, and (4) Onsite User Testing [36, 37]. The first two steps are covered in Study 1 while the final two steps are covered in Study 2. Our goal with this paper is to illustrate how to practically develop a connection between the algorithm developer and the algorithm agents, in order to create a feedback loop that encourages algorithmic accountability and agent empowerment.

4 STUDY 1: NEED FINDING

In Study 1, we conducted ethnographic field research and need-finding interviews to understand user needs and desires from the student assignment algorithmic mechanism. Given the novelty of the phenomenon, we employed a similar approach to Smyth et al. and avoided a hypothesis-oriented method of analysis, opting instead for inductive reasoning to identify key themes [57]. This approach is standard among several well-known qualitative analysis techniques.

4.1 Methods

Our goal in this ethnographic field research study was to understand the values and needs of the high school students when submitting their rank-order list of the 43 public universities to the Ethiopian Ministry of Science and Higher Education for the centralized university matching process. Within the context of algorithm design, we aimed to understand these values and needs so that we could convert them into variables that would be measured within our proposed algorithm. To achieve this goal, we conducted an ethnographic action research study that employs ICT research methodologies.

4.1.1 Data Collection. We began the study by conducting a set of formal, in-person, and semi-structured interviews (n=33). Data collection took place over a 4-week period in Addis Ababa, Ethiopia in the winter of 2022 by the first author. In this study, 27 of our 33 participants were 12th-grade secondary school students. The remaining 6 participants were public university students. Of our 33 participants, 18 identified themselves as female and 15 identified themselves as male. 17 of the students went to a public secondary school while the other 10 students went to a private secondary school. Both of these user subject groups were interviewed on the values and priorities which they believed would impact (or had impacted, if they were university students) their ranking of the 43 universities during the centralized university matching process. It was important for us to collect data that represented a diverse range of socioeconomic levels and a balanced gender distribution in order to have the closest possible representation of the millions of 12th grade students across the nation that undergo the student-university algorithmic mechanism in Ethiopia.

4.1.2 Data Analysis. To analyze the main themes discussed by the participants during this study, we transcribed data gathered over three weeks of interviews and labeled them accordingly using line-by-line open coding. We revised the labeling through a collaborative and iterative process, and then used axial coding to extract the relationship between themes. We identified five emerging themes that students wanted information on in order to rank their universities: (1) Infrastructure and Internet, (2) Available Departments, (3) Quality of Education, (4) Campus Environment, and (5) University Resources. We expanded on the subcategories for each of these emerging themes, as well as how they translate to our proposed algorithm parameters, in Table 1, 2, 3, 4, and 5. The proposed algorithm parameter for each subcategory represents the importance or relevance of that subcategory to the student's decision-making process when selecting a university. Each of these parameters will be assigned a weight and used to inform the proposed matching algorithm. Following our rigorous analysis of the interviews, we developed a conceptual framework to illustrate the Circles of Influence perceived by the typical student participant during the centralized university matching process in Ethiopia. The model shown in Figure 3 depicts the prevailing sentiment among the student participants. The sentiment that the students shared is that they lack agency in the selection of their assigned university, as they are insufficiently equipped with factual information about the 43 public universities in Ethiopia, which they are required to rank in order of preference. These key takeaways are supplemented by three findings.

4.2 Results

In this section, we present the findings from our formative qualitative work. In the Discussion section, we reflecting on these findings by offering two design implications which we employ for the development of our system in Study 2.

Subcategories	Description	Statement Example	Parameter
Building Infrastructure and Campus Internet Connectivity	The infrastructure of the campus buildings and the internet connectivity.	"I do not want to go to universities without adequate infrastructure and electricity. For example, my friend wanted to study computer engineering but she got into a university which doesn't even have steady electricity for lights, so she had to start working in the day and studying in the night classes."	$I_{infrastructure}$
Dormitory Sanitation and Food Quality	The quality of living facilities, particularly dormitories, and the quality of food options available to the students.	"One of my priorities is the food quality… Most people consider this because they are concerned about their health. If we are not offered good food options we will not be healthy so we will not able to study and we may fail our exams..."	I_{dorms}

Table 1. The subcategories, descriptions, statement examples, and resulting algorithm parameter I for the theme of **Infrastructure and Internet**.

4.2.1 *Qualitative Analysis.*

We present three main findings from this study. The first finding is that all of our student interviewees lacked confidence in making a rank-order list of university preferences beyond their top three choices, despite having to rank 43 universities. Hence, many of our students ranked the same well-known university as their first choice.

The next finding was that students were forced to trust and rely on word-of-mouth and social signaling subjective information from alumni with whom they had connections or family members. For example, when choosing which university to list as their first choice, students often deferred their decision to a family member who was an alumnus of a certain public university: "[M]y brother studied there and he told me to choose it because that is where those with top scores go" [A9].

The last finding was the most surprising for us: students did not have a single standardized, reliable, and factual information source to consult in order to learn about the attributes of the 43 universities they had to rank in their preferences for the university matching process. These 43 public universities had minimal to no online presence. Therefore, unless the high school provided an information packet to their students, which was a rare case and typically only mentioned within the private high schools that we visited, students were forced to rely on subjective and limited

Subcategories	Description	Statement Example	Parameter
Major Offerings	The different academic departments and fields of study offered by a university.	"If someone knows what they want to know beforehand, then they will target universities which teaches what they want. For example, in Ethiopia, Gondar University is said to be good for medicine studies, so it would be good if students who want to study medicine know this kind of information."	D_{fields}
Major Selection Process	The field selection process and acceptance rate at a university.	"I want to study 2 different fields (business management or geography). However, I can't make a choice. For geography, I have a lot of interest and I want to learn it. But, I don't have information on which universities it is taught at, what kind of courses it requires, or whether it aligns with my interests. So, if I got information on these kinds of things, it would really make the decision making process of which field I want to chose easier for me."	$D_{selection}$

Table 2. The subcategories, descriptions, statement examples, and resulting algorithm parameter D for the theme of **Available Departments**.

Subcategories	Description	Statement Example	Parameter
Quality of Pedagogy	Refers to the teaching methods, strategies, and the overall quality of education.	"I want to know the universities teaching strategy and quality of education."	$Q_{teaching}$
Practical Learning Opportunities for STEM Majors	The availability and quality of hands-on learning experiences for students studying science, technology, engineering, and mathematics (STEM) fields.	"Engineers should learn 75% of their material through practice and experience. However, at that university, engineering students learn 75% of their material through theory and 25% through experience, therefore, when they come out and get hired, they do not know anything and make mistakes."	$Q_{practice}$
Job Placement Rate	The likelihood of graduates finding employment in their desired field after graduation.	"The second reason why I chose these universities as my top three is because they have great post-graduate opportunities nearby. For example, if I were to go to Addis Ababa University and study properly, I can find a good job or I can become a lecturer there when I graduate. So, in other words, I would have good [job] opportunities."	$Q_{employability}$

Table 3. The subcategories, descriptions, statement examples, and resulting algorithm parameter Q for the theme of **Quality of Education**.

Subcategories	Description	Statement Example	Parameter
Peace, Safety, and Security	The peace, safety, and security of the campus environment.	"I will not rank universities that are in conflict zones higher on my list [...] I want to learn in peace and I want safety. I do not want conflict."	C_{safety}
Drug Prevention and Tolerance	The tolerance level for drug use on campus.	"I do not want to go to a university which tolerates drug addiction. At those kinds of universities, you could find gangsters and people like that, and then you don't know where you'll end up."	C_{drugs}
Distance from Part-Time Job Opportunities	The distance between the student's home and the university they are assigned to.	"I do not want to study in university because I want to work in the daytime and do night classes at a college. Even if I get into Addis Ababa University with my marks, I would still work in the day and study at night."	C_{jobs}
Distance from City and Landmarks	The distance of the student to the city and landmarks.	"I want to rank Adama University because I know the area of Adama well, it's not new to me, I can live there and I know the language of that area. And I say Hawassa because I'd love to go out in Hawassa, that's my dream."	$C_{recreation}$

Table 4. The subcategories, descriptions, statement examples, and resulting algorithm parameter C for the theme of **Campus Environment**.

Subcategories	Description	Statement Example	Parameter
Student Educational Resources	The availability of academic resources, including campus maps, libraries, laboratories.	"I wish I could see a map of all of the campus buildings and resources for the students like the offices, libraries, and laboratories."	$U_{learning}$
Student Extracurricular Opportunities	The range and quality of extracurricular activities offered by the university, including sports and clubs.	"For Bahir Dar . . . they have great extracurricular options. For example, I like football, and they have great clubs, I would love to play for their football club. So, Bahir Dar's resources, their atmosphere, and their campus' beauty are what draws me to that school."	$U_{extracurriculars}$
Student Mental Health Resources	The available mental health resources, such as counseling services for students struggling with mental health issues.	"I want to learn more about the healthcare and mental health resources at the universities. Because university life might disrupt our mental health, maybe, and it might challenge us because we are apart from our family so it will be important to know about our healthcare options."	U_{health}

Table 5. The subcategories, descriptions, statement examples, and resulting algorithm parameter U for the theme of **University Resources**.

information from their Circles of Influence to guide the development of their rank-order list. When questioned about their values and priorities that led them to pick their top three universities, several of the students stated that their top three choices were "totally based on rumors. I do not know [the validity] of any of these rumors" [A18]. **In other words, students were randomly ranking the majority of the 43 universities on their rank-order list.** From an algorithmic design perspective, this finding implies that regardless of the algorithm we develop, unless we address this concern, we will not be able to optimize the matches for neither the student nor the university side.

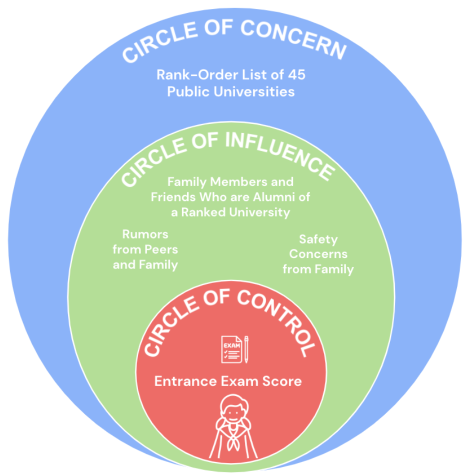


Fig. 3. Model of Student Circles of Influence on their Rank-Order List of 43 Public Universities.

4.2.2 *Quantitative Analysis.* The participants were questioned on their current process for finding information on the universities which they would need to rank. Of all of the secondary school students, 74% of them said they were going to rank Addis Ababa University as their number one choice and 55% said they were going to rank Hawassa University within their top three choices. 14% of the secondary school students were unsure of their top three choices. All of the public university participants, said that they ranked the university which they currently attend, Addis Ababa University, as their number one choice during the centralized university matching process. This statistic aligns with the fact that 74% of the secondary school participants said that they were going to rank Addis Ababa University as their first choice.

Of the 43 universities that they would be asked to rank during the centralized university assignment process, an overwhelming 92% of the secondary school students said that they knew how to rank their top three university choices and that they did not know the names of (or were going to randomly rank) the remaining 42 universities. This percentage includes both private and public secondary school students. This result was shocking as an important assumption of the algorithm behind the centralized university assignment is that the students know how to rank the 43 universities to accurately reflect their preferences. Even for their top three choices, many of the secondary school students stated that these top three choices were purely based on rumors.

5 STUDY 2: SYSTEM EVALUATION

In this study, we built and evaluated tools meant to meet the user needs and desires that we identified in our first study. This study validates our approach and findings from our first study.

5.1 Methods

In order to understand the impact of our novel assessment of fairness for student assignment algorithmic mechanisms, we conducted a within-subjects needs-based research study using a randomized controlled trial. We recruited n=40 students from various secondary schools in Ethiopia as our sample. In this study, we deployed a portal intervention that attempted to bridge the gap between algorithm developers and student needs. While using this portal during the intervention condition, students were exposed to comprehensive information about the universities, split up into categories that were directly informed from the results of student needs and values from Study 1, personalized acceptance rates to each university, and an AI-generated recommended rank-order lists of 43 public universities that attempted to help students fulfill the requirements of the rank-order list based preference language in student assignment algorithmic mechanisms [52]. We conducted quantitative data analysis in order to measure the performance of the proposed algorithmic mechanism and compare the student's satisfaction of their matching under our proposed algorithmic system versus their predicted matching under the current algorithmic system.

In order to understand the impact of providing high school students with information about the universities, along with personalized acceptance rates and an AI-generated recommended rank-order lists of 43 public universities on the fairness of the centralized university matching process in Ethiopia, this project follows the value-sensitive design methodology by considering the values that are important to the stakeholders involved in the university ranking process [20]. We investigate how the level of information on public universities, including students personalized acceptance rates to each one, affects their rankings, the study aims to uncover how to improve the university ranking process in a way that aligns with students values, needs, and aspirations, as well as those of the Ministry of Science and Higher Education in Ethiopia (MSHE). The use of machine learning to provide students with access to their personalized acceptance rates serves to give them a more personalized and user-centered approach to the university ranking process, aligning with the value of transparency. Through the discussion of the design implications and future work, we will continue to refer back to the values of the stakeholders. Our overarching goal in this work is redesigning a student assignment algorithmic mechanism that can be used to improve the university ranking process in a way that is ethical and responsible.

As mentioned above, we also developed a machine learning model which can predict a given student's acceptance rates to each university based on the current student-assignment algorithm. To develop this model, we used one-hot encoding and an ensemble of three different models and returned the class with the maximum number of votes. The three models we voted on were: a Random Forest Classifier model, a Logistic Regression model, and a Gradient Boosting Classifier model.

5.1.1 Data Collection. For the control condition, we provided a list of 43 universities in Ethiopia to the students and asked them to submit their initial rank-order list of the universities based on their preferences. Then, we distributed a survey to assess their ranking satisfaction using a 7-point Likert scale. Finally, we ran the Ethiopian Ministry of Science and Higher Education's (MSHE) current student-university assignment algorithm, asked them to submit their finalized rank-order list of the universities, and distributed another survey to assess their matched university satisfaction and trust in the MSHE's current algorithmic mechanism, again using a 7-point Likert scale.

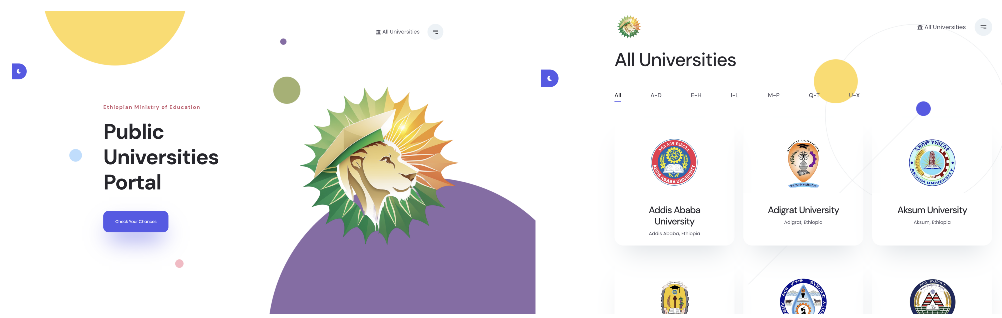
In the intervention condition, we exposed the students to a portal that provides them with comprehensive information about each university that they will be asked to rank. Then, we allowed students to rearrange the rank-order list according to their updated preferences as they went through the website and viewed in-depth details about each of the 43 universities. We also asked each of the students to click a button that would generate an AI-generated recommended rank-order list of universities. Upon clicking this button, the website also displayed where they would have been matched according to the two different student-university assignment algorithms (the MSHE's currently used one and our proposed one). At the end of the intervention, we asked the students to submit their finalized rank-order list of the universities and distributed a survey to assess the students' satisfaction of their matched university, their AI-generated recommended rank-order list, and their trust in our proposed algorithmic mechanism, using 7-point Likert scales.

5.1.2 *Data Analysis.*

For our data analysis, we performed a power analysis over a pilot study of 12 participants in order to determine our sample size. We first analyzed the data using descriptive statistics, including means and standard deviations. Then, we employed Kendall's Tau Distance, as a nonparametric measure of correlation between the rankings of the control and intervention groups, and a paired samples t-test, as an inferential static, to identify significant differences between the means of the two groups.

6 RESULTS

Our results show that over 60% of students found the intervention condition to have an impact on their rankings. In this section, we explain our evaluation of this system from our quantitative analysis.



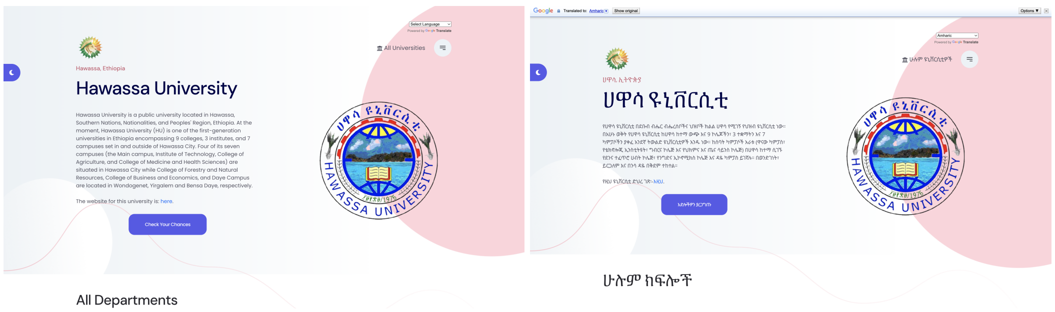
(a) (b)

Fig. 4. (a) Homepage of Ethiopian Public Universities Portal with AI Assistant. (b) Directory of all the universities on portal.

6.0.1 *Quantitative Analysis.*

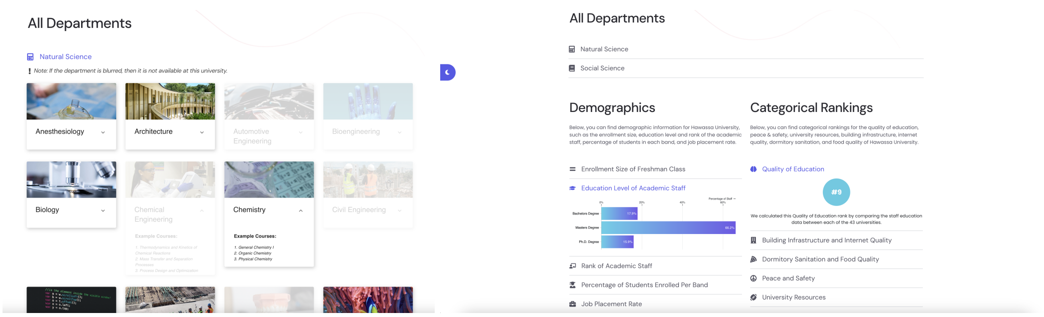
Descriptive statistics of the results for the feedback surveys revealed that the mean trust level and satisfaction level for their likely matched university of the students who used the website intervention were significantly higher than that of students who did not use the website. Students had a 28.3% increase in trust level between the current student assignment algorithmic mechanism and our proposed student assignment algorithmic mechanism. During the control condition, 58.9% of the students stated that they trusted the current student assignment algorithmic mechanism to have their best interests. However, during the intervention condition, 87.2% of the same students stated that they trusted the proposed student assignment algorithmic mechanism to have their best interests.

Inferential statistics through a paired samples t-test indicated that the Kendall's Tau difference in rank-order lists was greater for students after they went through intervention condition (Mdn = 10.45, SD = 17.26) than after they went



(a) (b)

Fig. 5. (a) Hawassa University Page (in English). (b) Hawassa University Page (in Amharic).



(a) (b)

Fig. 6. (a) All Departments for Hawassa University. (b) Demographics and Categorical Rankings for Hawassa University.

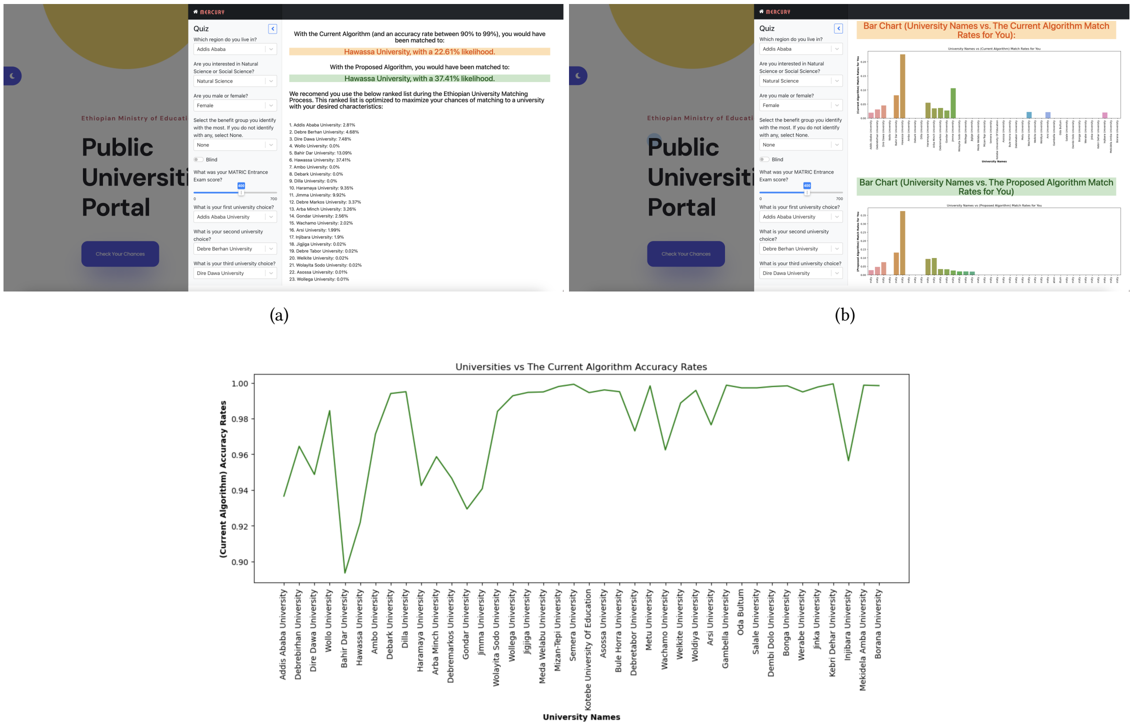
through the control condition (Mdn = 0, SD = 0), t(12) = 13, p = 0.05, d = 0.8563. This led to the results of our power analysis to be interpreted as sufficiently large in order for this study's results to be representative of the millions of students across the nation.

7 ETHOS: MULTI-CRITERIA STUDENT ASSIGNMENT MECHANISM

Ethos consists of three main technical components: 1) an information portal, 2) a machine learning-based system for suggesting universities, 3) a multi-criteria matching algorithm.

The proposed student assignment matching algorithm begins by defining the ministry's five separate rank-order lists of the 43 universities based on their individual score for the algorithmic parameters I, D, Q, C, and U that we defined from their respective, corresponding themes in Study 1: **Infrastructure and Internet**, **Available Departments**, **Quality of Education**, **Campus Environment**, and **University Resources**. In this way, our proposed algorithm models the choice problem for each school individually rather than employing the same choice rules for many schools [22]. For instance, Jimma University might be #3 in the **Quality of Education**-based rank-order list of the 43 universities, but #10 in the **Infrastructure and Internet**-based rank-order list of the 43 universities.

Then, the algorithm computes the weighted sum rank of each university, based on these characteristic-rankings. The weights would be determined by the appropriate clearinghouse, in this case the MSHE. Let's consider an example; let



(a) (b)

(c)

Fig. 7. (a) The model-generated recommended rank-order list for the given student data. (b) The predicted acceptance rates to each university for the current student based on the current algorithm and the proposed algorithm. (c) Our predicted acceptance rate model achieved a high accuracy rate for each university, ranging from 90% to 99%.

Jimma University have the following rankings: 20, 10, 3, 2, 1 for the five characteristics (with their respective weights): $I(0.1)$, D (0.2), $Q(0.3)$, $C(0.2)$, and $U(0.2)$. In this case, the weighted sum rank of Jimma Univeristy would be 20(0.1) + 10(0.2) + 3(0.3) + 2(0.2) + 1(0.2) = 5.5, which we would round up to 6. Once the algorithm has computed the weighted sum rank of each university, it generates a sorted list (in ascending order) of the universities based on their weighted sum rank. It breaks ties by prioritizing universities which have a higher **Quality of Education** ranking (this is because that was the characteristic that we identified was most desired by the students in Study 1).

At this point, the algorithm begins computing the student's acceptance rate for each university based on their top ten universities (which are a proxy for their most desired characteristics), their entrance exam score, and their demographic information. If the student's score and demographics satisfies a hard-to-reach quota requirement for the university, the matching algorithm adds a reward weight to that student's acceptance rate. Then, the algorithm will go through each university in the sorted list of weighted sum ranks of the universities. We want universities with the best characteristics to get their first pick of students. For each university that the algorithm goes through in the sorted list, the algorithm will sort the students in descending order based on their acceptance rates to that university. We also want students with the highest acceptance rates to that university to be chosen first. The algorithm iteratively admit students to that university until all of the available spots for each demographic group of that university have been filled. In other words, we have prioritized students who fulfill hard-to-reach quotas in this process by increasing their acceptance rate (with the reward weight that we mentioned in the prior paragraph). After the algorithm has filled all of the available spots for each demographic group at this university, it will repeat this process with the next university in the sorted list of weighted sum ranks. This algorithm terminates after all the universities have been checked.

The final output of the algorithm will be the matching of admitted students to each university for the 43 universities. While this algorithm does ensure an increased level of fairness in student assignment algorithmic mechanisms, it can also be integrated with any of the existing student placement mechanisms (TTC, Deferred Acceptance, etc.).

Algorithm 1 Multi-Criteria Matching Algorithm

Require: L_s: The student's rank-order list of universities; C_u The most desired characteristics by students (identified in Study 2); $L(u)$: The clearinghouse's rank-order lists of universities per characteristic u; $R(u)$: Reward function for each university c.

Ensure: Stable matching between students and universities that maximizes the students' chances of being assigned to a university with their most desired characteristics, while satisfying the universities' quotas and constraints.

$C(u_Y) \leftarrow$ empty list of weighted characteristic rank of the universities.
for each each university u_Y **do**
  $C(u_Y) \leftarrow \sum_{c \in C_u} w_c \cdot r(c, u_X)$, where w_c is the weight assigned to characteristic c, and $r(c, u_X)$ is the rank of university u_X for characteristic c in $L(u_X)$.
  $C(u_Y).append(u_Y)$.
end for
$SC(u_Y) \leftarrow$ sort($\{u_1, \ldots, u_N\}$, key = $\lambda u_Y : C(u_Y)$, reverse=True).
for each student s_X **do**
  $G(s_X), E(s_X), D(s_X) \leftarrow$ the given top ten universities (which are a proxy for their most desired characteristics after they have either viewed the website or the PDF), entrance exam score, and quota demographics for student s_X.
  $W_R(s_X) \leftarrow$ reward weights for hard-to-reach quotas of student s_X.
end for
for each student s_X and university u_Y in $L(s)$ **do**
  $A(u_Y, s_X) \leftarrow f(G(s_X, E(s_X), D(s_X), u_Y)$, where f maps the student's top ten universities, entrance exam score, and demographics to an acceptance rate for that university.
  if $len(W_R(s_X)) \geq 1$ **then**
    $A(u_Y, s_X) = A(u_Y, s_X)' \leftarrow$ acceptance rate adjusted upwards to satisfy desired hard-to-reach quotas.
  end if
end for
for each university u_Y in $SC(u_Y)$ **do**
  $T(u_Y) \leftarrow$ sort($\{s_1, \ldots, s_N\}$, key = $\lambda s_X : A(u_Y, s_X)$, reverse=True), where N is the number of students applying to university u_Y.
  $Q(u_Y) \leftarrow$ number of available spots for each demographic group.
  $A(u_Y) \leftarrow$ empty list of admitted students for university u_Y.
  for each student s_X in $T(u_Y)$ **do**
    if $A(u_Y)$ does not exceed the quota for any demographic group **then**
      $d \leftarrow$ demographic group of student s_X.
      if $Q(u_Y)[d] > 0$ **then**
        $A(u_Y).append(s_X)$.
        $Q(u_Y)[d] \leftarrow Q(u_Y)[d] - 1$.
      end if
    end if
  end for
end for
return the matching of students and universities in $A(u_Y)$.

8 DISCUSSION

In this paper, we took the case of student university assignments in Ethiopia as a case study to implement an end-to-end participatory algorithm design process. The result was Ethos, our proposed student assignment algorithmic mechanism, which uses deliberative democracy [5, 8] to translate the collective needs and values of stakeholders into algorithm parameters that increase mechanism fairness and empower stakeholders. Ethos also holds each of the universities in the system accountable by customizing the choice problem for each school rather than applying the same choice rules for many schools [21].

Challenging the ecosystem the algorithm will live in is essential and one of the core findings of this paper. In the case of this research, this meant questioning and iterating on the preference elicitation process during data collection. We had to slow down the algorithm design process in order to stop and listen to the participants. This led to a bottom-up approach during our algorithm design process which included traveling overseas in Study 1 in order to conduct the ethnographic research. This slowed down approach to participatory algorithm design was pivotal in helping us to identify the needs and desires of the agents controlling the algorithm output: the students.

In Study 2, we found that the level of information available on public universities had a significant effect on students rankings. Specifically, students who used the website intervention showed higher satisfaction levels with their university matches and trust in the centralized university process than those who did not use the website. An advantage of this study was that we were able to assess several different dimensions of student assignment and matching under a controlled setting.

8.1 Design stakeholder input into the algorithm design process.

A core value in this work is showing how, as algorithmic mechanism design researchers, we must resist techno-determinism [33] and increase stakeholder empowerment by integrating deliberative democracy [9] and value-sensitive design [20] into our algorithm design process. This requires awareness of participatory mechanisms and developing methods to expand participatory methods to incorporate complex algorithm design challenges. A central aspect of this work will be to allow the people affected by the system to audit and modify the algorithm by designing human-centered feedback and equity into the mechanism from the beginning rather than trying to patch up a broken system with an algorithm. We followed this principle in Study 1, when we took in student feedback on values to consider for the resulting novel value-centered matching algorithm that we tested with students in Study 2.

8.2 Build in community.

Leading with grounded theory in our research meant constantly and intentionally questioning the ecology that the algorithm would exist in and being in community with the participants in order to incorporate their participation in the proposed algorithm's research and design. This was prior to any data analysis or software development. The extra time and deliberate intentionally we took to understand the user perspectives early on in our research, during Study 1, was necessary and led our research to better suit them, which we witnessed in Study 2. Our goal with this paper to create an algorithm which encourages feedback loops, redistributes the power imbalance between the algorithm and the agent, and ultimately, gives agency back to the community.

8.3 Limitations

One limitation of Study 1 is that we only recruited participants from one geographic area in Ethiopia (Addis Ababa). Despite our diverse set of socioeconomic levels, there is still a stark and disproportionate level of privilege that students in the capital city have comparison to the lesser developed and rural regions of Ethiopia. A limitation of Study 2 is that it had low ecological validity because the controlled conditions in this experiment are not necessarily what happen in the real world.

8.4 Design Implications

From Study 1, we present the following two design implications for future research in student assignment participatory algorithm design: (1) in order for participants to be able to make informed rankings, they need accessible information, and (2) in order for universities to get diverse student bodies, they need to explain to the students how their university adheres to the student's needs, aspirations, and desires.

Following the results of Study 2, we learn that students consider a portal and predicted acceptance rate model as necessary tools for any student assignment mechanism because they empowered them when making their rank-order lists of universities. In this study, students described how these "transformative" tools provided them with crucial information about their chances of acceptance which would help to comfort them wherever they matched, instill their confidence in the matching process, and ultimately contribute to their happiness and success.

9 CONCLUSION

Our paper is a real-world research application of enacting theories of justice in the algorithm design process by gathering the needs, aspirations, and desires of the people affected by the algorithm. Our research utilizes participatory design approaches for the development of an algorithmic system for school assignment in Ethiopia. Through ethnographic research with the people who are subject to decisions by the system, we developed a theory of fairness in this system. We used these insights to develop and evaluate a technical system based on the desires and values of our participants. This project exemplifies an effort to amplify the voices of participants to lead the direction of the algorithmic research.

ACKNOWLEDGMENTS

We express our gratitude to the participants in both of our studies for their willingness to share their experiences and valuable insights.

REFERENCES

[1] Atila Abdulkadiroglu, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. 2017. *Minimizing justified envy in school choice: the design of New Orleans' OneApp.* Technical Report. National Bureau of Economic Research.

[2] Atila Abdulkadiroğlu and Tayfun Sönmez. 2003. School choice: A mechanism design approach. *American economic review* 93, 3 (2003), 729–747.

[3] Atila Abdulkadiroğlu, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. 2020. Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights* 2, 4 (2020), 425–442.

[4] Nikhil Agarwal. 2015. An empirical model of the medical match. *American Economic Review* 105, 7 (2015), 1939–1978.

[5] Ari-Veikko Anttiroiko. 2003. Building strong e-democracy: the role of technology in developing democracy for the information age. *Commun. ACM* 46, 9 (2003), 121–128.

[6] Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. 2010. *Handbook of social choice and welfare.* Vol. 2. Elsevier.

[7] Michel Balinski and Tayfun Sönmez. 1999. A tale of two mechanisms: student placement. *Journal of Economic theory* 84, 1 (1999), 73–94.

[8] Benjamin Barber. 2003. *Strong democracy: Participatory politics for a new age.* Univ of California Press.

[9] Derek WM Barker. 2017. Deliberative justice and collective identity: A virtues-centered perspective. *Political Theory* 45, 1 (2017), 116–136.

[10] Anna Maria Barry-Jester, Ben Casselman, and Dana Goldstein. 2015. The New Science of Sentencing. *The Marshall Project* (2015). https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing

[11] Ruha Benjamin. 2019. Assessing risk, automating racism. *Science* 366, 6464 (2019), 421–422.

[12] Ruha Benjamin. 2019. *Captivating technology: Race, carceral technoscience, and liberatory imagination in everyday life.* Duke University Press.

[13] Ruha Benjamin. 2023. Quantified Racism. *The Quantified Worker: Law and Technology in the Modern Workplace* (2023), 293.

[14] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–12.

[15] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society* 3, 1 (2016), 2053951715622512.

[16] Yves Cabannes. 2004. Participatory budgeting: a significant contribution to participatory democracy. *Environment and urbanization* 16, 1 (2004), 27–46.

[17] Robert Chambers. 1995. Poverty and livelihoods: whose reality counts? *Environment and urbanization* 7, 1 (1995), 173–204.

[18] Yan Chen and Tayfun Sönmez. 2006. School choice: an experimental study. *Journal of Economic theory* 127, 1 (2006), 202–231.

[19] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'connell, Terrance Gray, F Maxwell Harper, and Haiyi Zhu. 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.

[20] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

[21] Battal Doğan. 2019. Mechanism Design Approach to School Choice: One Versus Many. In *The Future of Economic Design: The Continuing Development of a Field as Envisioned by Its Researchers*. Springer, 469–472.

[22] Federico Echenique and M Bumin Yenmez. 2015. How to control controlled school choice. *American Economic Review* 105, 8 (2015), 2679–2694.

[23] Haluk Ergin and Tayfun Sönmez. 2003. Games of school choice among boston students. *Manuscript. MIT and Koç University* (2003).

[24] Florian Evequoz, Johan Rochel, Vijay Keswani, and L Elisa Celis. 2022. Diverse Representation via Computational Participatory Elections-Lessons from a Case Study. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–11.

[25] Gregory Ewing and Ibrahim Demir. 2021. An ethical decision-making framework with serious gaming: a smart water case study on flooding. *Journal of Hydroinformatics* 23, 3 (2021), 466–482.

[26] Martin Fisher. 2006. Income is development: KickStart's pumps help Kenyan farmers transition to a cash economy. *Innovations: Technology, Governance, Globalization* 1, 1 (2006), 9–30.

[27] Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P Dickerson, and Vincent Conitzer. 2020. Adapting a kidney exchange algorithm to align with human values. *Artificial Intelligence* 283 (2020), 103261.

[28] David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[29] Marzyeh Ghassemi, Tristan Naumann, Peter Schulam, Andrew L Beam, Irene Y Chen, and Rajesh Ranganath. 2020. A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings* 2020 (2020), 191.

[30] Steven Glazerman and Robert H Meyer. 1994. Public school choice in Minneapolis, midwest approaches to school reform. In *Proceedings of a Conference Held at the Federal Reserve Bank of Chicago*. 26–27.

[31] Rustamdjan Hakimov and Onur Kesten. 2018. The equitable top trading cycles mechanism for school choice. *International Economic Review* 59, 4 (2018), 2219–2258.

[32] Matt Haney, Stevon Cook, and Rachel Norton. 2018. Developing a Community Based Student Assignment System for SFUSD. *San Francisco Unified School District. Retrieved April* 13, 2020 (2018), 2019–20.

[33] Robert L Heilbroner. 1967. Do machines make history? *Technology and culture* 8, 3 (1967), 335–345.

[34] Yoan Hermstrüwer. 2019. Transparency and fairness in school choice mechanisms. *MPI Collective Goods Discussion Paper* 2019/11 (2019).

[35] Zoë Hitzig. 2020. The normative gap: mechanism design and ideal theories of justice. *Economics & Philosophy* 36, 3 (2020), 407–434.

[36] Melissa R Ho, Emmanuel K Owusu, and Paul M Aoki. 2009. Claim Mobile: Engaging conflicting stakeholder requirements in healthcare in Uganda. In *2009 International conference on information and communication technologies and development (ICTD)*. IEEE, 35–45.

[37] Melissa R Ho, Thomas N Smyth, Matthew Kam, and Andy Dearden. 2009. Human-computer interaction for development: The past, present, and future. *Information Technologies & International Development* 5, 4 (2009), pp–1.

[38] Caroline M Johnston, Simon Blessenohl, and Phebe Vayanos. 2020. Preference elicitation and aggregation to aid with patient triage during the covid-19 pandemic. In *Workshop on Participatory Approaches to Machine Learning*.

[39] Anson Kahng, Min Kyung Lee, Ritesh Noothigattu, Ariel Procaccia, and Christos-Alexandros Psomas. 2019. Statistical foundations of virtual democracy. In *International Conference on Machine Learning*. PMLR, 3173–3182.

[40] Onur Kesten. 2004. Student placement to public schools in the US: Two new solutions. *Unpublished mimeo, University of Rochester* 9 (2004).

[41] Onur Kesten. 2012. On two kinds of manipulation for school choice problems. *Economic Theory* 51 (2012), 677–693.

[42] Fuhito Kojima, M Utku Ünver, et al. 2010. The" Boston" School-Choice Mechanism. (2010).

[43] Amanda Kube, Sanmay Das, Patrick J Fowler, and Yevgeniy Vorobeychik. 2022. Just resource allocation? How algorithmic predictions and human notions of justice interact. In *Proceedings of the 23rd ACM Conference on Economics and Computation*. 1184–1242.

[44] Heidi Ledford. 2019. Millions affected by racial bias in health-care algorithm. *Nature* 574, 31 (2019), 2.

[45] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.

[46] Min Kyung Lee, Ishan Nigam, Angie Zhang, Joel Afriyie, Zhizhen Qin, and Sicun Gao. 2021. Participatory algorithmic management: Elicitation methods for worker well-being models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 715–726.

[47] Cecilia Machado and Christiane Szerman. 2016. Centralized admission and the student-college match. (2016).

[48] Thayer Morrill. 2015. Two simple variations of top trading cycles. *Economic Theory* 60, 1 (2015), 123–140.

[49] Ritesh Noothigattu, Snehalkumar Gaikwad, Edmond Awad, Sohan Dsouza, Iyad Rahwan, Pradeep Ravikumar, and Ariel Procaccia. 2018. A voting-based system for ethical decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[50] Parag A Pathak. 2011. The mechanism design approach to student assignment. *Annu. Rev. Econ.* 3, 1 (2011), 513–536.

[51] Parag A Pathak. 2017. What really matters in designing school choice mechanisms. *Advances in Economics and Econometrics* 1, 12 (2017), 176–214.

[52] Samantha Robertson, Tonya Nguyen, Cathy Hu, Catherine Albiston, Afshin Nikzad, and Niloufar Salehi. 2023. Expressiveness, Cost, and Collectivism: How the Design of Preference Languages Shapes Participation in Algorithmic Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[53] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. 2021. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[54] Alvin E Roth and Elliott Peranson. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American economic review* 89, 4 (1999), 748–780.

[55] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2020. A human-centered review of algorithms used within the US child welfare system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[56] Lloyd Shapley and Herbert Scarf. 1974. On cores and indivisibility. *Journal of mathematical economics* 1, 1 (1974), 23–37.

[57] Thomas N Smyth, Satish Kumar, Indrani Medhi, and Kentaro Toyama. 2010. Where there's a will there's a way: mobile media sharing in urban india. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 753–762.

[58] Tayfun Sönmez and M Utku Ünver. 2011. Matching, allocation, and exchange of discrete resources. In *Handbook of social Economics*. Vol. 1. Elsevier, 781–852.

[59] Halil Toros and Daniel Flaming. 2018. Prioritizing homeless assistance using predictive algorithms: an evidence-based approach. *Cityscape* 20, 1 (2018), 117–146.

[60] Kentaro Toyama. 2015. *Geek heresy: Rescuing social change from the cult of technology*. PublicAffairs.

[61] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[62] Mohammad Yaghini, Andreas Krause, and Hoda Heidari. 2021. A human-in-the-loop framework to construct context-aware mathematical notions of outcome fairness. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 1023–1033.

[63] Takuya Yokota and Yuri Nakao. 2022. Toward a decision process of the best machine learning model for multi-stakeholders: a crowdsourcing survey method. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 245–254.

[64] Angie Zhang, Alexander Boltz, Jonathan Lynn, Chun-Wei Wang, and Min Kyung Lee. 2023. Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[65] Angie Zhang, Olympia Walker, Kaci Nguyen, Jiajun Dai, Anqing Chen, and Min Kyung Lee. 2023. Deliberating with AI: Improving Decision-Making for the Future through Participatory AI Design and Stakeholder Deliberation. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.

[66] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric PS Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.