No Fairness without Awareness

A statistical study into equity at THUAS.

let's change
YOU. US. THE WORLD.

Project description and justification, v. 1.0.3

Learning Technology & Analytics Lectorate | 07-04-2024

DE HAAGSE HOGESCHOOL

Table of contents

1 Introduction

The project No Fairness without Awareness of the lectorate Learning Technology & Analytics (LTA) aims to map equal opportunities for students within The Hague University of Applied Sciences (The HHs) and - where necessary - develop advice to improve them. Analysis of inflow, throughput, and outflow (*student journeys*) is essential to gain insight into the position and meaning of our college for students and the region.

1.1 What do programs and students gain from this research project?

Students get answers through these analyses as to whether De HHs offers them equal opportunities for admission and obtaining a Degree in the foreseeable future. They also gain insight into the possible bottlenecks and can then question their program or the support provided by De HHs. It also becomes clear to them their chances of getting a job or continuing their studies at the level they want after graduating from De HHs. These insights come from analyzing the programs and researching the student journey of groups of students. An example of the latter is research into the transition of students with prior education on the Caribbean islands to The Hague and their student journey.

Programs can use these analyses to improve their educational policies or teaching. For example, a program with insight into the development of intake over the past ten years can be used to identify possible areas for improvement to keep the program relevant. Or we show the impact of a change in the curriculum on the composition of new intake and the effectiveness of this change. With partners such as the Central Bureau of Statistics (CBS), we can, for example, analyze the graduates' success in the labor market.

Any bottlenecks in the student journey are thus identified, and study programs can use these insights to eliminate them.

1.2 What does THUAS gain from this research project?

This research project contributes to THUAS's institutional plan, 'Inquiry-based learning with impact.' In particular, regarding the strategic themes, I. Quality of education and research (ambition 1), and IV. An inclusive community (ambitions 11 and 12).

> **AMBITION 1**
>
> **Continuous improvement of the quality of education.**
>
> We strive for the highest possible quality of our education to maximise the impact and value for students and the professional field. Quality is also essential for preventing unnecessary study loss. We will use a short-cycle and results-oriented approach, utilising study data for the further

> development of feasible and teachable curricula.
>
> **AMBITION 11**
>
> **Wellbeing of students and staff come first.**
>
> We provide an environment in which the well-being of students and staff is paramount, so that everyone can study and work at THUAS with pleasure and pride. In doing so, we are mindful of the individual backgrounds, needs, and ambitions of the members of our community. Our community offers togetherness, trust, and responsibility - strong motivating factors that enable our students and staff to contribute to our community and society as a whole.
>
> **AMBITION 12**
>
> **Inclusive culture.**
>
> We value a diversity of perspectives, and we break through exclusionary mechanisms. This requires openness and courage to speak up when discrimination occurs, even when it is unintentional, and we recognise that it is everyone's responsibility to do so. We are mindful that students and staff have different starting positions and may therefore encounter different bottlenecks. It is our responsibility to identify these bottlenecks and to work with students and staff to remove them or provide alternative routes.

In addition, it contributes to THUAS' goals of inclusivity as articulated in the vision document "An inclusive college: collaboration for inclusive education and research," dated November 26, 2021, version 1.0.

> **The vision of diversity and inclusion**
>
> (…) An inclusive college is an environment that provides equal opportunities for all, values equality, and thereby ensures physical, social, cultural, educational, and work-related accessibility for students and staff from diverse backgrounds. In an environment like this, close attention is paid to preventing obstacles or forms of exclusion and discrimination. This is an area of shared responsibility for everyone in the college.
>
> **Principle of anti-discrimination**
>
> (…) Access to our college (and thus access to education) is guaranteed to the maximum extent possible. Faculties and departments take the necessary measures to remove barriers and to be critical at all times of implicit and explicit criteria, rules, and procedures that (unintentionally) lead to exclusion and discrimination. The university monitors and quickly takes corrective action when policy principles, procedures, work instructions, and practices (unintentionally) lead to exclusion and discrimination.

1.3 Position of the project in the research lines of the lectorate, the knowledge center, and THUAS

The project contributes to the lectorate's lines of research in Student Analytics, Institutional Analytics, and Inclusion Analytics. Not only with new insights but also with the method we are developing to arrive at those insights. Additional research questions we ask in this regard are: which reporting forms provide immediate insight, and which take more effort? What triggers action or not? What leads to impact, and what does not? How can we best serve different target groups in this? These questions are not addressed in this project description.

The research further falls within the Transformative Technology research line of the Global & Inclusive Learning Knowledge Center. It aligns with THUAS's Digital Future and Equitable Society research themes, because of the intersection Machine Learning and equal opportunity.

1.4 Purpose and structure of this document

This paper describes the further substantiation of the importance of the project (chapter 2), the operationalization and research questions of the project (chapter 3), the sources we use for this purpose (chapter 4), the process for delivery and editing (chapter 5), the methods of analysis (chapter 6), the expected results (chapter 7) and the reproducibility of the research (chapter 8).

2 Substantiation of interest

This chapter provides a theoretical and ethical rationale for the importance of the research project.

2.1 Theoretical foundation

The research program No Fairness without Awareness aims to identify equal opportunities for students within THUAS and - where necessary - provide advice on improving them.

It is a given that differences exist among students and affect their entry, progression, exit from higher education, and subsequent success in the labor market. Where this reflects differences in student abilities, this is only sometimes an immediate problem. But where they do not, there may be a lack of equal opportunity. This is the subject of study of the research project "No Fairness without Awareness."

Compared to other European countries, Dutch education has a unique structure in secondary education. Differentiation in students' study paths is already made at 12, whereas in other European countries, this is at 15. This fosters inequality (Copier, 2022).

The term "equal opportunity" is widely used, but it is essential to define its content further. For equal opportunity in higher education, three approaches scale over time and scope (Elffers, 2022): 1) equal opportunities to realize learning potential, 2) equal learning and graduation opportunities for equal potential, and 3) equal opportunities to find a good place in society for different talents. For details, see Figure 1.

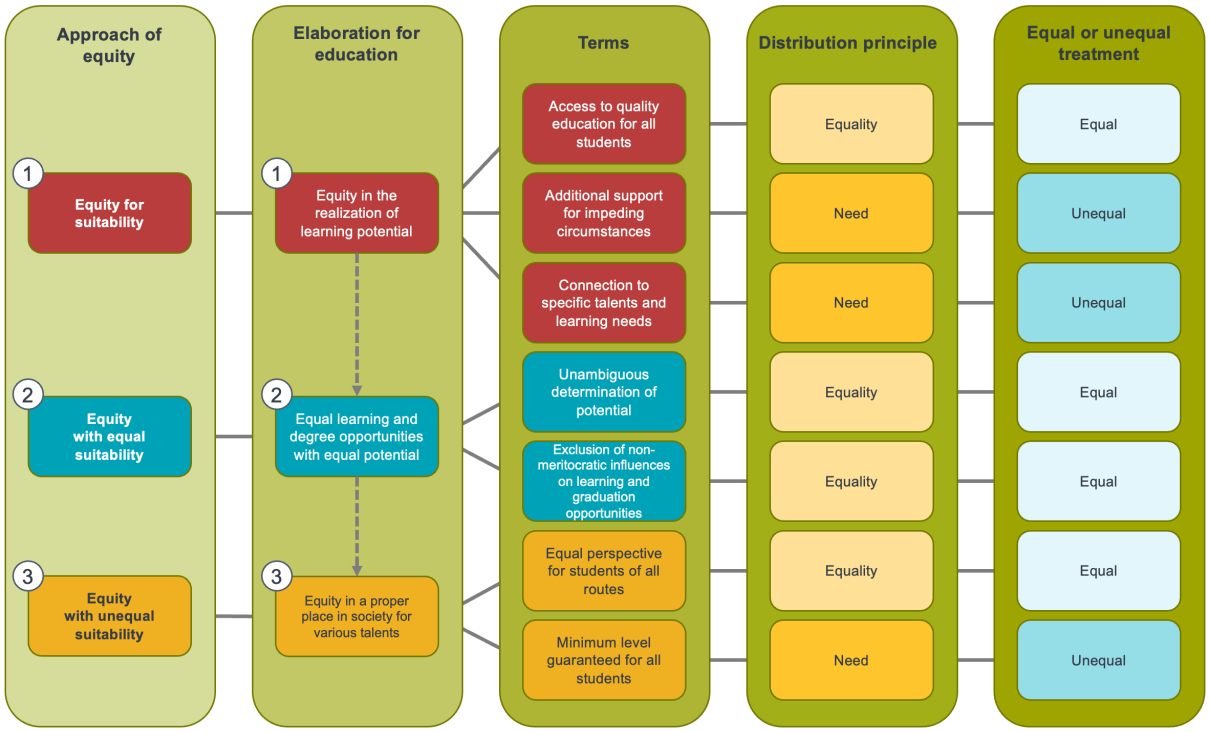

Figure 1: Equity Framework - Elffers (2022)

In English, there is a difference between the terms *equality* and *equity* (Espinoza, 2007). While "equality" focuses on group equity as a distributive principle, "equity" focuses on individual needs. In this paper and research, we choose the first principle, group equity, as our lens. In the second approach - equal opportunity for equal ability - a further distinction can be made (Espinoza, 2007), p. 347:

1. **Survival** (*educational attainment*) - The likelihood that students from different social groups will remain in the school system up to a certain level. This could include the propaedeutic level or degree.
2. **Outcome** (*educational achievement based on test performance*) - The probability that students from different social groups will learn the same things at the same level at some point in the school system. Here, one can think of the outcome at the subject level.

We focus on "survival" in this research, taking different moments of the student journey as reference points c.q. bottlenecks. For an elaboration, see section 3.2: Equal learning and graduation opportunities at equal potential or equal chances for successful progression and outflow.

2.2 Ethical considerations

Based on the principle of academic integrity, this document was prepared to be transparent in the research and accountable for the data collected and methods applied. In this section, we outline

the ethical considerations that were made. Appendix 1 used the deon checklist for data scientists to address a wide range of ethical questions.

The research aims to contribute to the solution of a social and ethical issue: unequal opportunities in higher education.

The research is aimed at answering an ethical question: to what extent do students at THUAS have equal opportunities in their studies and - upon completion - in further studies or on the job market? The goal is for THUAS to be a safe and stimulating environment for all students in the region. In doing so, we align ourselves with the agenda of the City of The Hague's Equal Opportunities Alliance (2019-2021):

> All children and youth are entitled to equal opportunities in education. It is important that the talents of all children and young people are used to their full potential. This is of great importance not only for the future of children and young people themselves but also for the future of our society.

To investigate equal opportunity, we study the background characteristics of students that may affect equal opportunity.

To answer the research question, an estimate of students' starting situations is needed. These include meritocratic, such as previous achievement and prior education, and non-meritocratic characteristics, such as socioeconomic status, gender, and age. In part, these are directly related to the student - such as gender, age, and prior education. In part, they are indirectly derived - such as socioeconomic status based on the student's geographic origin at the neighborhood and district level; without these data, we cannot estimate differences in socioeconomic status. To conduct analyses, the modeling uses neighborhood-level SES-WOA scores from CBS.

Including these factors is not uncontroversial. Even though they are not special personal data and the data are anonymized, some say it is better not to include these data in analyses to avoid bias (Pedreshi et al., 2008). However, this is naive (Hardt et al., 2016): removing these variables does not rule out the possibility that these characteristics may still be present as proxies through other variables in the dataset and still affect the outcomes of an analysis or derived decision. We want to detect possible bias in existing data to prevent it from becoming a rule in the development of statistics in higher education, which may result in institutional discrimination (Pedreshi et al., 2008). Here applies: "Fairness through awareness" (Dwork et al., 2011). This requires examining data from recruitment, admission, teaching, and testing processes at THUAS to see if there may be inequality of opportunity traceable to these characteristics.

We increase understanding of research to date on equity in education.

An exploratory study of equity in education identified 30 factors in five levels: student (9), family (6), school (7), neighborhood (5), and society (3) (Badou & Day, 2021). Not all factors translate into available data, so we include a select number of them in this study, particularly at the student, school, and

district levels. One of the recommendations from the study was follow-up research on the mutual interaction and ranking of these factors. To illustrate, gender and age may interact in study success after one year, but it may turn out that, age is more important than gender. The statistical research we propose using Machine Learning contributes to this, as it can examine both interaction effects and the relative contribution of each factor to a forecasting model.

Equal opportunity does not equal discrimination.

The distinction between equal opportunity and discrimination is important. The research aims to examine equal opportunity in the context of higher education and THUAS University of Applied Sciences in particular; this can be distinct from discrimination. Discrimination is the willful making of distinctions based on one or more non-meritocratic background characteristics of a person to treat them adversely and unequally. In this study, equal opportunity is a neutral term that states whether a student in transition in the student journey can expect equal outcomes based on equal merit. By this, we mean personal abilities, such as intelligence, commitment, executive skills, and motivation.

In presenting research findings, we take into account the avoidance of disclosure risk and stigmatization.

The research itself must not encourage discrimination or stigmatization. Therefore, possible risks of disclosure and reliability are taken into account derived from the CBS guidelines:

- A minimum number of 10 observations per cell is used for tables and graphs.
- Percentages are published only when the denominator contains at least 100 observations.
- Visualizations are aggregated to higher levels (e.g., by cohort, education, neighborhood).

In the application of Machine Learning in this research, we apply the ethical principles of the EU.

The EU has established Ethical Guidelines for Trustworthy AI (April 8, 2019). These contain four ethical guidelines to which we are committed in this research:

- **Respect for human autonomy** - The Machine Learning algorithms from this research will not be applied for practical applications, such as a prediction model for individual student advising, but rather to “enhance, complement and strengthen human cognitive, social and cultural skills.

> The fundamental rights on which the EU is founded aim to ensure respect for people’s freedom and autonomy. People working with AI systems must be able to retain their full and effective self-determination and participate in the democratic process. AI systems should not unjustly subjugate, coerce, mislead, manipulate, condition, or drive people. Instead, they should be designed to augment, complement, and enhance human cognitive, social, and cultural abilities. The allocation of functions between humans and AI systems must follow human-centered design principles and leave open meaningful opportunities for human choice. Thus, human supervision and human control of work processes in AI systems must be provided.

- **Prevention of damage** - The Machine Learning from this research will be applied precisely to determine where the risk for negative impact on minorities may be. During the research (Sub-research II), we will involve students in developing the analysis on bottlenecks, which ML will be used for. We are exploring the possibilities with the Inclusion Office and the Partner Up! program.

> AI systems must not cause or increase harm or otherwise adversely affect people. This means protecting the dignity as well as the mental and physical integrity of people. AI systems and the environment in which they operate must be safe and secure. They must be technically robust and care must be taken to ensure that they do not provide room for malicious use. Vulnerable people should be given more attention and should be involved in the development and installation of AI systems. Specific attention should be paid to situations where AI systems may cause or increase negative impacts due to inequality in terms of power or disposition of information, for example, between employers and employees, between businesses and consumers, or between governments and citizens. Prevention of harm also means considering the natural environment and all living things.

- **Justice** - The research was established to promote equity and equal opportunity. Machine Learning is being used to discover biases in historical data from THUAS in order to deduce what bias should be avoided in future regulations or applications of Machine Learning.

> The development, installation, and use of AI systems must be equitable. We recognize that there are many different interpretations of equity, but are convinced that equity has both a substantive and a procedural dimension. The substantive dimension implies a commitment to ensure the equal and equitable distribution of both benefits and costs and to ensure that individuals and groups are free from unjust bias, discrimination and stigmatization. If unjust bias can be avoided, AI systems could even enhance social justice. Equal opportunity in terms of access to education, goods, services and technology should also be promoted. In addition, the use of AI systems should never have the effect of misleading (end) users or restricting their freedom of choice. Further, equity implies that AI professionals must respect the principle of proportionality between means and ends and carefully consider how to balance competing interests and objectives. The procedural dimension of justice includes the ability to challenge and effectively appeal decisions made by AI systems and by the people who run them. To do so, the entity responsible for the decision must be identifiable and the decision-making process must be explicable.

- **Accountability** - In addition to this paper's justification for the research project, the models and input factors used will be described for justification, and the source code of these models will be publicly available.

> Accountability is crucial for creating and maintaining users' trust in AI systems. This means that processes must be transparent, the capabilities and purpose of AI systems must be made openly known, and decisions must be explainable - to the extent possible - to those directly or indirectly affected by them. Without that information, a decision cannot be properly challenged. It is not always possible to explain why a model produced a particular outcome or decision (and what combination of input factors contributed to it). These cases are called "black box" algorithms and require special attention. In these situations, other accountability measures (such as traceability, auditability, and transparent communication about the system's capabilities) may be necessary, provided that the system as a whole respects fundamental rights. The extent to which accountability is needed depends heavily on the context and the severity of the consequences, should the result be incorrect or otherwise inaccurate.

In presenting the outcomes, we use different perspectives and elaborate on the advantages and disadvantages.

There is no single definition of fairness; one can and should look at it in multiple ways. We will introduce those nuances in presenting results and discussions with stakeholders to avoid one dominant, oversimplified view of fairness in education, policy, or guidance emerging within THUAS. We will present the different perspectives side by side with advantages and disadvantages. In addition, we will include limitations of the research, such as relational fairness (Fish & Stark, 2022), that cannot be answered by the method used.

We give students a voice in our understanding of fairness and the possible translations of insights into applications in THUAS.

We want to avoid that the research results in the translation to practice can lead to a narrowing of images about groups of students.

According to THUAS's inclusion principles of 'inclusive manners' and 'inclusive governance,' we do not want to reduce students to a category and encourage active participation to improve educational policy and practice. We, therefore, discuss research findings with students through various forums and channels. Possible student networks are faculty councils, the Inclusion Office, the Knowledge Center Global & Inclusive Learning networks, and students from the CMD program.

We will test with these students beforehand the images of fairness and selection of data from the research proposal, and as the research progresses, the outcomes and communication about them. We will ask students in the CMD program to examine the visualizations of the software we will use for usability and usability for students. We will also guide faculty on communicating the possible use of the insights to their students and staff we are testing with students.

3 Operationalization of approaches to equity of opportunity

To identify equal opportunities and possible bottlenecks, we examine Elffers' three approaches to equal opportunities based on 1) origin and entry, 2) progression and exit with or without a Degree, and 3) success in further education or the labor market.

3.1 Equal opportunities to realize learning potential or equal opportunities for entry

This first perspective implies both equal and unequal treatment:

> "This first perspective implies both equal and unequal treatment:"This requires, first and foremost, that all learners have equal access to education of sufficient quality to enable them to develop their sight, which amounts to equal treatment. But it equally requires that adjustments be made when students face circumstances that may limit the realization of their potential and that customization be provided to students with diverse talents and learning needs. That precisely requires unequal treatment." (Elffers, 2022)

For this purpose, we examine our students' **origin, intake,** and **likelihood of entering** THUAS. The assumption is that all students from the supply area of THUAS are equally likely to be educated at THUAS. The expectation is that this is unlikely to be the case. Possible bottlenecks may arise for several reasons (Elffers, 2022): a) no access to a good education for all students, b) too little support for impeding circumstances, and c) no connection to specific talents and learning needs.

Table 1: Operationalization equal opportunity to realize learning potential

Condition	Operationalization	Possible dates
a) No access to a good education for all students	1. Probability of inflow based on demographic and socioeconomic background.	Gender, age, socioeconomic background, date of application
	2. Selective recruitment by THUAS.	THUAS recruitment criteria.
	3. Selection of students in selective programs	Selection Criteria
b) Insufficient support for impeding circumstances.	1. Supporting students with disabilities	Development of volume and nature of applications from students with disabilities

Condition	Operationalization	Possible dates
	2. Support for students with non-regular prior education (state examination or 21+)	Development of amount and nature of intake of students with non-regular prior education
	3. Support for students with deficiencies	Criteria for granting facilities
c) Failure to match specific talents and learning needs	1. Selective orientation and study choice by students.	Probability of inflow from the region plotted against actual inflow
	2. Quality of connection depends on prior education (havo, mbo, foreign Degree, etc.).	Prior education, country of highest prior education, type of affiliation
	3. Quality of affiliation depends on type of affiliation (directly after prior education, gap year, switch)	Connection and intake of Caribbean students

The **research questions** we want to explore in this perspective based on THUAS's study data are:

1. To what extent do prospective students have equal or unequal opportunities for admission to education at THUAS?
2. Are students treated equally or unequally in the admissions process by THUAS where appropriate - given their opportunities?

3.2 Equal learning and graduation opportunities at equal potential or equal chances for successful progression and outflow

The second perspective calls for equal treatment of students at equal potential.

> This requires, on the one hand, an unambiguous determination of students' potential and, on the other hand, the exclusion, as far as possible, of the influence of non-meritocratic factors on students' chances of learning and obtaining a Degree in a particular form of education. (Elffers, 2022)

For this purpose, we examine students' **progression** through their studies and **outflow with or without a degree**. The assumption is that, for equal ability, students are equally likely to progress by subject, stage, or for the entire study and equally likely to obtain a degree or drop out after each year of study.

Table 2: Operationalization equal learning and graduation opportunities with equal potential

Condition	Operationalization	Possible dates
a) Unambiguous determination of potential	1. Quality of education and testing	Background and prior education characteristics as in perspective 1
b) Exclusion of non-meritocratic learning and graduation opportunities at equal potential	1. Opportunity for advancement and graduation	Throughput in courses, transition points (such as propaedeutic or BSA, subsequent stages of study), and graduation within a foreseeable time (nominal study time or nominal study time + 1 year)
	2. Probability of dropout	Dropout after each academic year
	3. Opportunity to take elective courses	Accessibility of desired education
	4. Opportunity for facilities for students with disabilities	Availability of facilities for taking education or tests
	5. Opportunity to pursue alternative teaching and testing formats for achieving learning outcomes.	Availability of alternative forms for achieving learning outcomes

The **research questions** we want to explore in this perspective based on THUAS's study data are:

1. To what extent do students have equal or unequal opportunities for advancement in education by THUAS?
2. Are students treated equally or unequally in advancement in education by THUAS where appropriate - given their opportunities?

3.3 Equal opportunities for different talents to find a good place in society or equal opportunities for further study or position in the labor market

This third perspective implies both equal and unequal treatment:

> This makes demands, first and foremost, on the perspectives offered by different routes in education. The routes must offer an equal perspective of achieving a good place in society. (…) Regardless of the degree of inequality in pay, safeguarding the chances of a good life for all requires, in any case, that education manages to bring all pupils in all routes up to a level that is the minimum necessary for full participation in education, in the labor market and in society at large. Given pupils' the varying levels and learning needs, achieving that minimum level requires unequal treatment. (Elffers, 2022)

To this end, we examine students' **success** in **further education** or the **labor market**. The assumption is that under unequal eligibility, students are equally likely to pursue further education or find a job at the level, within the graduation domain, and within the foreseeable time.

Table 3: Operationalization equal opportunity for different talents to have a good place in society

Condition	Operationalization	Possible dates
a) Equal perspective for students of all routes	1. Opportunity to pursue further study after the various forms of education and training provided by THUAS	Enrollment types (full-time, part-time, dual) and program types (bachelor's, ad, master's) 1CHO data on transfer to an advanced program
	2. Job opportunities after the various forms of education and training at THUAS.	Enrollment types (full-time, part-time, dual) and course types (bachelor, ad, master) CBS microdata on labor market success
b) Minimum level guaranteed for all students	*[to be determined]*	*[to be determined]*

The **research questions** we want to explore in this perspective based on the study data from THUAS are:

1. To what extent do students have equal or unequal chances of advancement to an advanced degree or level job within the graduation domain and in the foreseeable future after education by THUAS?
2. Are students treated equally or unequally in advancement after education by THUAS where - given their opportunities - this is necessary?

3.4 Delineation

For all formal KPIs and reports of THUAS, these cannot be derived from the research of the lectorate but can only be based on the management dashboards and reports of the departments of Education, Knowledge & Communication (OKC) and Business & Control (B&C).

Because it appears from the operationalization that we are using Dutch sources (DUO, CBS) for these data, this study is, in principle, limited to the intake of students with a Dutch prior secondary education who enter a bachelor or associate degree program at THUAS, full-time, part-time or dual.

To determine whether there are (significant) differences with students with an international prior education, those enrolments will be included in exploratory analyses.

For now, the research is being conducted for the first time; whether it will be repeated more often and with what frequency is currently unknown and will depend on the outcomes, needs of THUAS, and the development of the research program of the professorship.

3.5 Connection to existing research within THUAS.

Definitions will be made explicit and compared with the formal definitions used within OKC and B&C reports. The study may deviate from definitions to investigate more dynamics in the student population, such as dropouts before October 1, or because data is not (yet) available, such as 'nationality.' In all these cases, deviations are substantiated and reported.

Where possible or desired, we compare outcomes with OKC or VH surveys, such as the HBO Monitor (conducted by ROA of Maastricht University and DESAN Research Solutions).

4 Data sources

Depending on the equal opportunity perspective, we examine different sources:

Table 4: Sources by perspective

Perspective	Phase	Resources
1. Equal opportunity to realize learning potential	Orientation and entry	**THUAS: OKC** - study data on applications and enrollments from the Study Data Project **THUAS: OKC** - CRM data on recruitment and orientation **CBS: Open data** on demographic development and socioeconomic status of neighborhoods and districts in the Netherlands **OCW: Open data** on students and graduates in secondary education
2. Equal learning and graduation opportunities for equal potential	Advancement and obtaining a degree	**THUAS: OKC** - study data on study progress, dropout, and graduation from the Study Data Project
3. Equal opportunity for different talents to have a good place in society	Outflow to further study or labor market	**CBS: Microdata** on labor market success **1CHO: Study data** on progression to further education

4.1 Explanation of edits by data source

The basis for the study is the dataset from OKC's study data project. We call this dataset "the study data analysis set."

4.1.1 Basic details of THUAS.

The **THUAS study data analysis set**:

- contains a set of variables per student per year of enrollment for the 2012 through 2022 cohorts (see Appendix 2 - Variables_Analysis Set_202306068 for a description of these variables);
- is provided anonymously upon request by OKC's IR & Analytics team is thus a secondary source:
  - excludes data such as student number, name, place of birth, date of birth, or other directly identifiable characteristics;

  - does not contain any special personal data.

4.1.2 Enrichment based on additional data sources

The LTA lectorate enriches the study data analysis set into a tertiary source from additional, secondary sources for a better understanding of the student's context and the program the student is following: THUAS' Customer Relationship Management (CRM) and data from the Association of Universities of Applied Sciences (VH), the Education Executive Agency (DUO), the Central Bureau of Statistics (CBS), Studychoice123, OpenStreetmaps (OSM) / OpenTripPlanner (OTP), and 1 figure HO (1CHO; also from DUO).

- The **THUAS CRM data**:
  - contain data by program of The HHs on the numbers of visitors to orientation activities and conversion rates from orientation to enrollment;
  - are aggregated by program level by college year;
  - are provided upon request by the Marketing & CRM team of the Marketing & Communications unit of the OKC department of The HHs;
  - are available from college year 2022-2023;
  - are linked by program level based on the name of the program.
- The **VH study program data**:
  - contain national, public data on first-year intake, progression, and outflow of students in programs at colleges;
  - are aggregated by level of education by college year;
  - are publicly available through the VH website via the dashboards influx, enrollments, and degrees (cohorts 2017 to 2022) en student retention, drop out, and switch (cohorts 2017 to 2020);
  - are linked by educational level based on the study program name.
- The **DUO open educational data**:
  - contain, in the form of open education data, public, nationwide data on (a) (expected) enrollment, progression, and outflow of secondary and higher education schools (cohorts 2017 through 2022; projections: 2022 through 2041) and (b) establishment addresses;
  - are aggregated by school type and educational institution by college year;
  - are publicly available through the website of DUO;
  - are linked at the school level based on the BRIN6 code: school and establishment.
- The **CBS open data**:
  - include CBS public statistics from Statline by neighborhood and district level;

  - kerncijfers wijken en buurten 2004-2022 worders used by neighborhood and district of characteristics concerning population, housing, energy, education, labor, income, social security, and facilities as proxies of these characteristics of the student at the start of the study;
  - sociaal-economische status; scores per wijk en buurt, regio-indeling 2021 are used by neighborhood and district as a proxy for socioeconomic status[1] of the student at the start of the study;
  - for linkage at the neighborhood and district level[2] the 4-digit zip code of the residential address at the time of graduation from the student's prior education is aggregated to a neighborhood and district code within a municipal code[3].

- The **Studiekeuze123 NSE data**:

  - are the NSS benchmark file: a multi-year file that contains student assessments and so-called indicator data (such as contact time);
  - are used to benchmark differences in student opinions from the National Student Survey by program;
  - are made available upon request to interested parties, including researchers;
  - are linked by program level based on the program's name.

- The **OSM** and **OTP** data:

  - are used to calculate the public travel times between the 4-digit zip code of the residential address at the time of graduation from prior education and a possible subsequent education at alternative educational institutions with similar educational offerings and, therefore, the likelihood of entering The HHs;
  - are linked by student based on the 4-digit zip code of the residential address at the time of graduation.

- The **DUO 1CHO** data:

  - are supplied to De HHs by DUO;
  - are used to analyze progression to subsequent study by program;
  - contain all participation and results of students in higher education from 1991 through October 1 of the previous year;
  - are linked by program level based on the name of the program.

[1] CBS: "De sociaal-economische status (SES-WOA) van gemeenten, wijken en buurten in Nederland. Deze status wordt beschreven in termen van de financiële welvaart, het opleidingsniveau en het recente arbeidsverleden van particuliere huishoudens op 1 januari van het verslagjaar."

[2] CBS: "Onderdeel van een gemeente, bestaande uit één of meerdere buurten. Vaak komt een wijk overeen met een woonplaats of een deel van een grotere woonplaats."

[3] Daar waar een postcode in meerdere buurten, wijken of gemeenten ligt, wordt de postcode gekozen met het hoogste aantal woningen, conform de methode die het CBS hier zelf voor hanteert.

In addition to enriching existing resources, the professorship - in collaboration with OKC's IR&A study data team and faculty staff - will develop a primary resource, the **study programs dataset**. This dataset:

- contains specific features per program by cohort, such as formal name, popular name, BSA limit, application deadlines, entry requirements for parts of the curriculum, grade scales, tracks, etc.
- is built fundamentally from the **study data analysis set** - as many characteristics as possible are taken from available data or derived;
- is linked by education level based on the name of the education.

4.1.3 Rationale for processing by data source

Table 5: Reasoning for processing by source

Data source	Data	Rationale
Study data analysis set (THUAS) = HHs-01	Basic data per student per study per year on personal demographic characteristics, orientation, prior education, application, admission, enrollment, study progress, dropout, or study success	Research on student journeys: equal opportunity at potential bottlenecks (inflow, progression, outflow)
CRM (THUAS) = HHs-02	Participation in the orientation of THUAS, conversion from orientation to enrollment by degree program	Researching student journeys: equal opportunity at potential bottlenecks (intake)
VH = VH-01 + VH-02	National overview with equivalent programs concerning 1) enrollment, enrolments, and degree, and 2) study success, dropout and study change by program	Benchmark research on national developments per course to distinguish between trends and exceptions (inflow, transfer, outflow)

Data source	Data	Rationale
DUO = DUO-01 + DUO-02	Basic data by school on 1) (expected) intake and 2) flow and outflow from secondary and higher education schools	Examining the extent to which THUAS's student population reflects the region based on market shares and proportionality of enrollment in THUAS programs (intake)
DUO = DUO-03	Branch addresses by school	Examining the distance from these schools to THUAS branches (inflow).
DUO = DUO-04	1CHO - Student participation and outcomes in higher education	Research on student journeys: research on progression to further study (outflow)
CBS = CBS-01	Characteristics concerning population, housing, energy, education, labor, income, social security, and services	Research on student journeys: equal opportunity at potential bottlenecks (inflow, progression, outflow)
CBS = CSB-02	Socioeconomic status	Research on student journeys: equal opportunity at potential bottlenecks (inflow, progression, outflow)
Studiekeuze123 = SK123-01	Student ratings and so-called interpretation data (such as contact time)	Research on student journeys: calculating the probability of choosing a course of study at THUAS (intake)
Travel times dataset (LTA) = LTA-01	Travel times from PC4 codes	Research on student journeys: calculating the probability of choosing a course of study at THUAS (intake)
Study programs dataset (LTA) = LTA-02	Characteristics by program by cohort	Research on student journeys: equal opportunity at potential bottlenecks (inflow, progression, outflow)

Data source	Data	Rationale
Study choice probability (LTA) = LTA-03	The probability of choosing a course from THUAS based on PC4	Research on student journeys: equal opportunity at potential bottlenecks (inflow, progression, outflow)

4.1.4 Schematic representation of data sources and links

In summary, the data sources and links are arranged as follows:

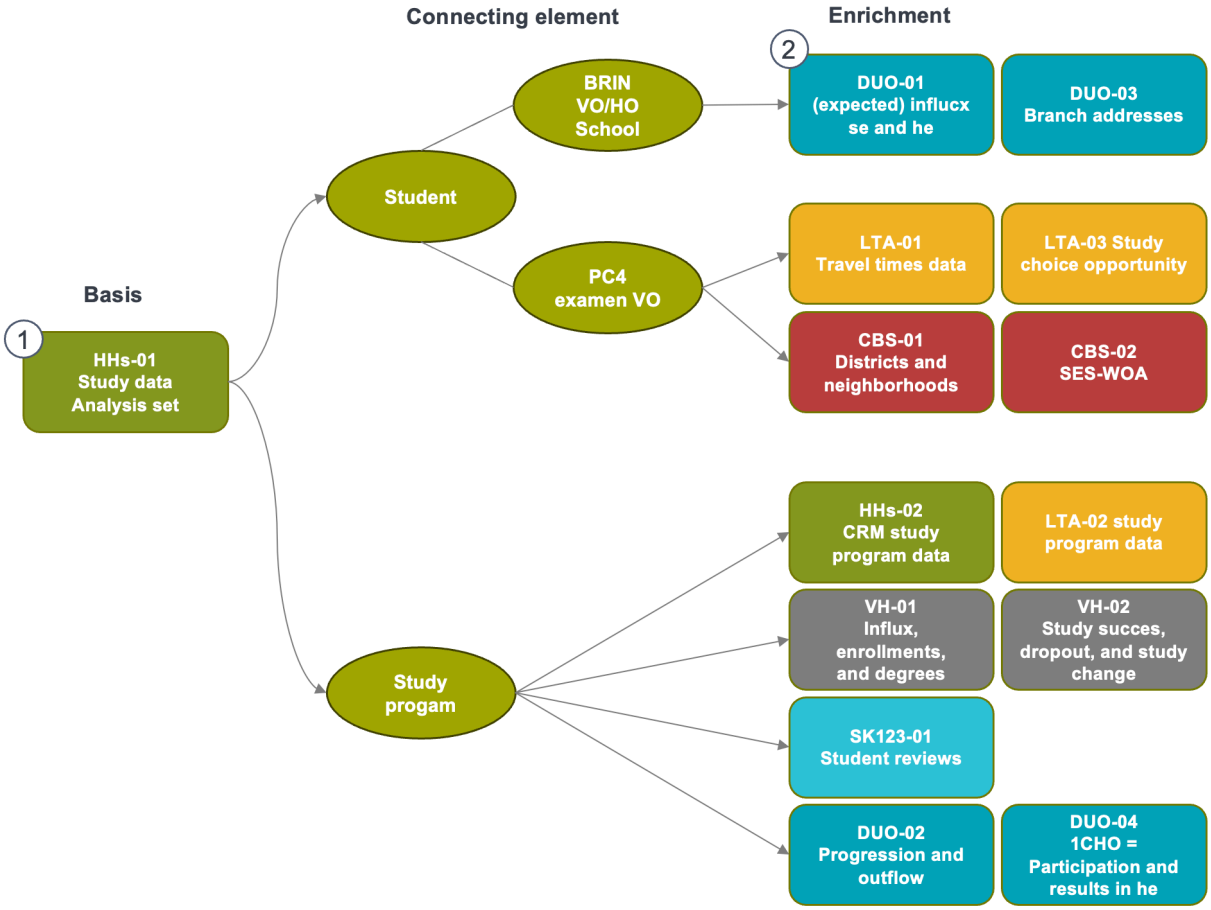

Figure 2: Data sources and links

5 Process of delivery and processing

5.1 Application to deliver or download

- Data from THUAS is provided by the respective teams based on a formal request. This document is intended to substantiate and document this process.
- Non-public data from public institutions is requested separately according to the procedures set up for this purpose; for the time being, this only concerns Stichting Studiekeuze123.
- All other data is publicly available and is downloaded via public access.

5.2 Data management

- The data is stored on the lecturer's Research Drive in accordance with the procedure for scientific data management established by the college library of THUAS.

5.3 Editing and enrichment

- The data is transformed into research results in several steps:
  - **documentation** of the resource in accordance with *THUAS's Guide to Folder Structure and Data Documentation* (version 1.0, February 2021, University Library).
  - **reading** in the study data analysis set;
  - **cleaning up** the dataset: unifying names and fields where necessary and filling missing values where necessary (e.g., with the mean or median of a field);
  - **saving** to an .rds and .fst file format;
  - **linking** with additional sources based on PC4, program name and cohort;
  - **enrich** the dataset: develop additional fields not yet available but derivable, such as "number of days of enrollment after the application deadline," "travel time to alternative study at another educational institution.
- Each step in the operation is additionally documented, and each intermediate result is saved as a separate file so operations can be easily traced afterward and reported in an account or publication.

6 Methods of analysis

In this section, we describe the analysis method for each approach to equity of opportunity.

1. An **analysis of (possible) study choices** to determine which (groups of) students are proportionally more or less represented at THUAS than we would expect based on the geographical location of the branches of THUAS. This will answer the question of what bias there is in the intake at THUAS.
2. An **analysis of bottlenecks and equal opportunities in the student journeys** of programs at THUAS relative to different groups of students. This will answer the question of the bias in student journeys at THUAS.
3. An **analysis of success in further education or the labor market** to determine which groups of students have more or fewer opportunities after college. This will answer the question of what bias exists in students' post-graduation success at THUAS.

For a visual elaboration, see Figure 3.

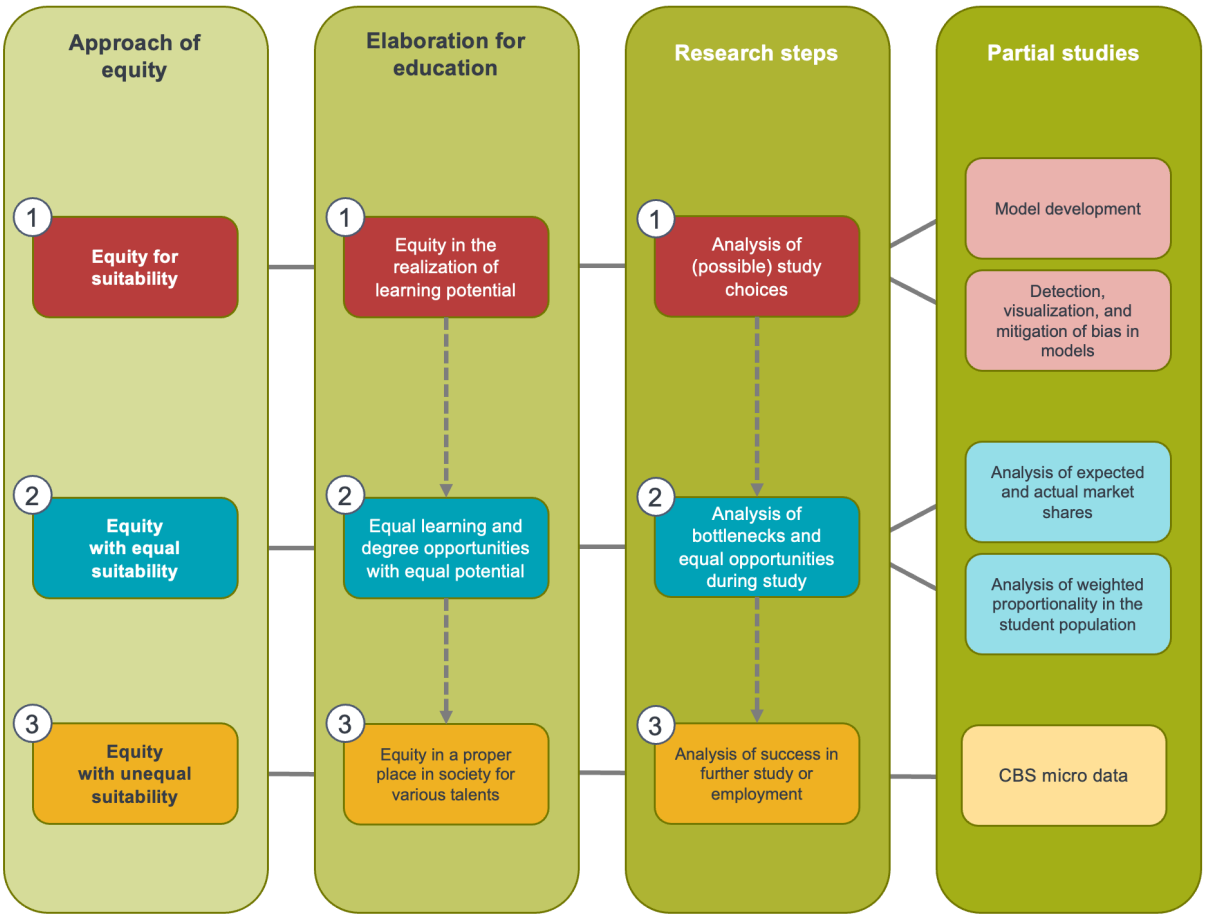

Figure 3: Partial studies

6.1 Partial study I - Analysis of (possible) study choices

This substudy answers the research questions from the first part of the operationalization of Equal opportunities to realize learning potential or equal opportunities for entry.

6.1.1 Analysis of expected and actual market shares

The first step of the substudy consists of calculating study choice probabilities of the intake of THUAS relative to the spreading area or region. For this purpose, we use a Huff Gravity model (Huff, 1963) using the REAT package and QGis. This model is used in retail when calculating the probability that customers will choose a store. As a store gets further away, the customer's willingness to travel there decreases (*distance decay*). However, some factors dampen this decay: good accessibility by public transport or private transport (*time-space compression*) and the attractiveness of the store, for example, the size or range of the store. As an illustration, see Figure 4.

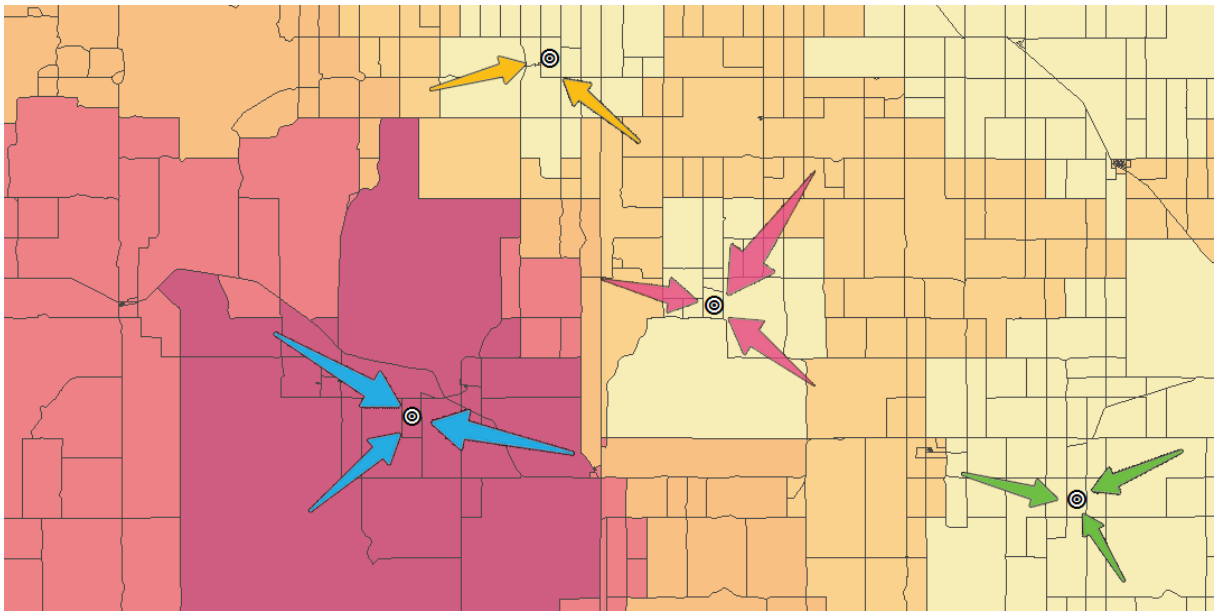

Figure 4: Example of a Huff Gravity Model (source: GISGeography)

We apply this concept and model to students' choice of one of the educational institutions in the Hague region.

- The assumption is that students will also choose an educational institution with the fewest barriers to pursuing their desired studies.
- Aspects conducive to removing barriers are limited travel time, a good reputation, and limited entrance requirements as dampening factors.
- Since public transportation is free for students, the cost of public transportation is not included.

- Furthermore, we assume that most students in their first year of study still live at home or that it is equally difficult to find a room in all cities, so available housing can be excluded as a factor.

For this analysis, four datasets are edited or developed and analyzed:

- The **historical outflow** data of secondary schools (havo and mbo) within a 90-minute travel time radius during rush hour on a Monday from each branch of THUAS (sources: DUO-02 + DUO-03 + LTA-01). The assumption is that students are willing to travel a maximum of 90 minutes to THUAS. The market share THUAS has in this outflow is then calculated.
- For each program, it is then determined what **alternative equivalent programs** are offered at other institutions within a 2x 90-minute radius. In theory, these are the equivalent programs that a student can choose instead of a program at THUAS; these programs are geographically located in North Holland, South Holland, Zeeland, Utrecht, and Brabant. For each of these programs, NSE scores are collected over the period 2010 to 2023 (source: SK123-01).
- For each enrollment, **travel times** are calculated in the first year at the program of THUAS and equivalent programs from PC4 onwards at the time of obtaining the prior degree (sources: LTA-01 and LTA-02).
- A **study choice probability** (a weighted probability of enrollment) is then calculated and save, based on travel times, reputation and entrance requirements per program (source: LTA-03). The weighting of the components herein remains to be determined.

6.1.2 Analysis of weighted proportionality in the student population

After calculating the expected and actual market shares by neighborhood, we can calculate the possible **inflow bias** based on weighting by neighborhood based on the probability calculation from the earlier Huff analysis. The bias works both ways: students from areas with lower odds count more heavily, while those with higher odds count less heavily. By normalizing the distribution of probabilities, we can project the weighting around the number 1, keeping the total population the same. In this way, we have developed a **propensity score**[4] for entry to THUAS and **balance** the overall student population.

We then compare **the weighted population to the unweighted population** on background characteristics, such as prior education, age, gender, and socioeconomic status. This allows us to know the **background bias** for the total population and have a better view of minorities and majorities.

We conduct this analysis by site and course based on one benchmark year to be determined.

[4] A propensity score is a score that expresses an item's probability of falling into a particular category. In this analysis, it is propensity to study at one of THUAS' sites.

6.2 Partial study II - Analysis of bottlenecks and equal opportunities during the study

This partial study answers the research questions from the second part of the operationalization of Equal learning and graduation opportunities at equal potential or equal chances for successful progression and outflow, in other words: equal opportunities in the student journey. The data are analyzed according to the 'Cross-industry Standard Process for Data Mining' (CRISP-DM) methodology (Chapman et al., 2000), differentiated according to the 'Exploratory Model Analysis' method (Biecek & Burzykowski, 2021) with the associated DALEX package and the fairmodel package. These packages are part of DrWhy.AI, an ecosystem of packages or analyses for explainable and fair AI.

First, one or more general models are developed, after which they are checked and corrected for fairness. This approach provides insight into the extent to which different groups have equal opportunities or a lack thereof in THUAS and which factors need to be "turned" to counteract unequal opportunities.

6.2.1 Model development

Three conditions are important in the development of a prediction model:

1. **Validation of prediction.** For each model prediction, we must be able to verify how strong the evidence is that supports the prediction.
2. **Justification of prediction.** For each model prediction, we must be able to explain which variables affect the prediction and to what extent.
3. **Speculation of prediction.** For model prediction, we must be able to explain how the prediction would change if the values of the variables in the model changed.

The development is an iterative process that is repeated several times based on the outcomes in the different steps as the understanding of the data grows (see Figure 5). The DALEX package helps to develop models transparently, maximizing the interpretability of the outcomes.

Table 6: Methods of analysis

Steps	Parts	Explanation
1. Data preparation	a. Research design	This document and, in particular, this chapter.
	b. Data collection c. Data cleaning	The collection and processing of data as described in Data sources.
2. Data understanding	a. Data exploration	Exploration of basic features based on frequency counts in table form, graphs, or maps.

Steps	Parts	Explanation
		Calculation and visualization of statistical relationships, considering the statistical composition of the population (distributions) and associated assumptions, and corrections for multiple testing. Examination of missing values, outliers (*outliers*) and imbalances in distributions (skewness or disproportionality).
	b. Feature selection	Choice of variables in which the differences between groups is significant or of general interest from the literature.
	c. Feature engineering	The following static operations take place: • Padding of missing values • Removal of outliers • Correction of skewness or disproportionality using transformations
3. Model assembly	a. Model selection b. Parameter estimation c. Hyper parameter tuning	Development of explainable Machine Learning models for predicting transitions in study: entry/selection, progression in the study along different points (e.g., BSA, propaedeutic year, additional parts of curricula with entry requirements such as internship, exchange with a foreign country, graduation project), and dropout or degree. Testing a range of models depending on the outcome variable (binary - e.g., pass or fail - or continuous - e.g., number of EC or grade level). Possible models include linear regression (GAM/GLM), classification and regression trees (CART), random forest (RF), stochastic gradient boosting, and bagged CART. Depending on the type of model, additional tuning of the settings (hyperparameters) is required (e.g., in RF). These can be searched and selected automatically.

Steps	Parts	Explanation
4. Model audit	a. Data validation b. Model validation c. Model benchmarking	Examination of the model on: • The various predictive components • A sensitivity analysis for the predictions • The predictive power: the final selection will be based on a comparison of the performance of these models, based on standard performance measures such as contingency tables for classifications (e.g., male/female versus degree yes/no) or the Reciever Operating Characteristic (ROC) and the Area Under the Curve (AUC) for analyses on rank orders (probabilities based on multiple variables). • The variable importance by explanatory variable provides insight into differences between programs and groups of students, which provide insight into the underlying mechanisms (Shmueli, 2010). • The effect of the explanatory variables on the predictions; this includes what-if scenarios via Ceteris Paribus analyses, which can indicate how a prediction changes if a characteristic were slightly different, e.g., the age of a student • Any residuals (unexplained effects)
5. Model delivery	a. Model rollout b. Documentation c. Communication	The results can be shared through the publishing tools provided for these models (model studio and arena). The use of these tools will be determined once the research has reached this stage and depends on the degree of aggregation of the data.

6.2.2 Detection, visualization, and mitigation of bias in models

After one or more models are developed according to the above method in the first phase, in the second phase, we investigate the possible bias in these models using the `fairmodels` package

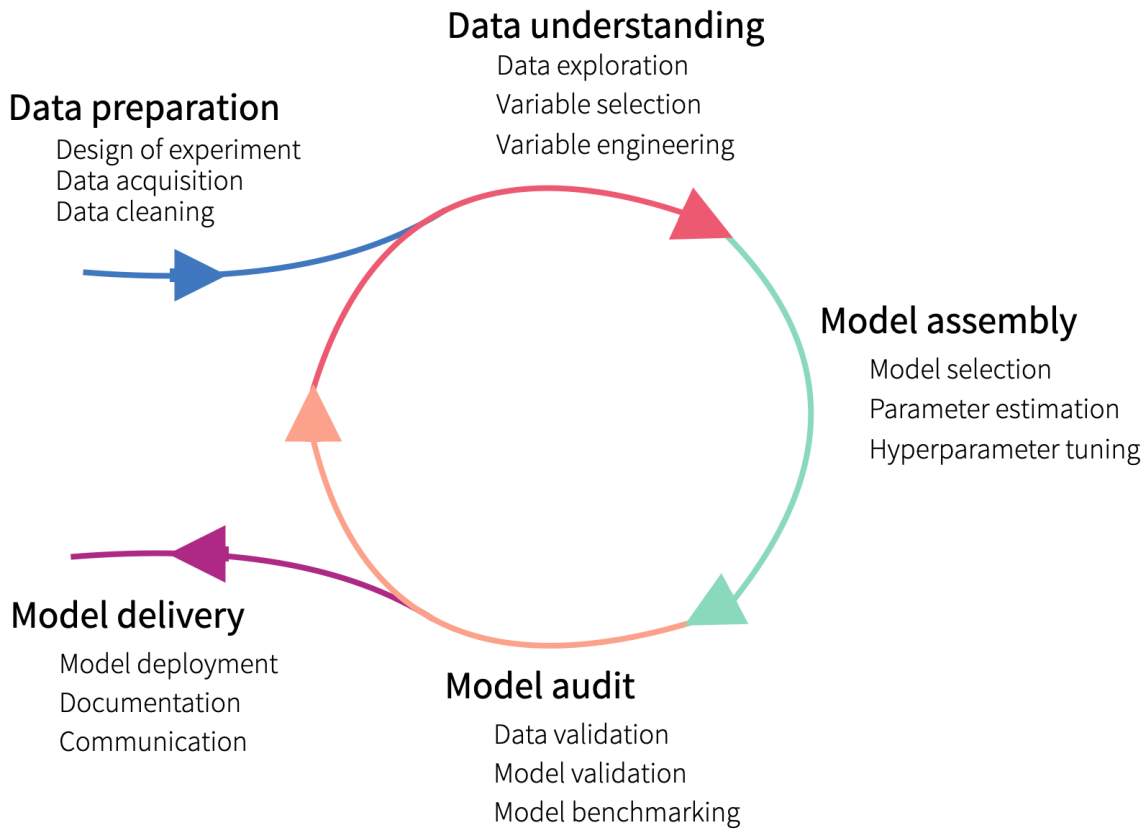

Figure 5: CRISP-DM methodology; visualization of Biecek and Burzykowski (2021)

(Wiśniewski & Biecek, 2022). We focus on binary classifications, e.g., "a student did or did not pass the BSA. Here are four main questions (Wiśniewski & Biecek, 2022): How do we measure bias? How do we detect bias? How do we visualize bias? How do we counter bias?

Establishing criteria for bias

There are three criteria for measuring bias or fairness: independence, separation, and sufficiency (Barocas et al., 2019). To illustrate, we explain these concepts using the achievement of a BSA of 50 EC for males and females.

1. **Independence** - The independence criterion means that the probability of obtaining 50 EC in year 1 (the BSA limit) should be equal between men and women. This is fair from a social perspective.
2. **Separation** - When assigning a negative BSA, there may be a misallocation: a female student may get a negative BSA because of an exam that has not yet been marked. The separation criterion implies that the ratio of rightly versus wrongly a negative BSA should be equal in the subgroups (men and women). This is fair from the student's perspective.

3. **Sufficiency** - The sufficiency criterion means that the ratio of rightly versus wrongly a negative or positive BSA should be equal in subgroups (men and women). This is fair from the organization's perspective (in this example, the educational institution).

A problem and challenge is that these criteria can only partially be valid simultaneously (Barocas et al., 2019). This requires balancing the different criteria or giving preference to one of them.

Detecting and visualizing bias with a cutoff criterion and fairness checks

It is common in Machine Learning to assume the 4/5 rule when balancing[5]: discrimination occurs when the selection of members of a minority group is less than 80% of members of the most selected group (*Code of Federal Regulations. Section 4d, uniform guidelines on employee selection procedures (1978)*, 1978). The `fairmodels` package, therefore uses a margin of 0.8. This is a *disparate impact analysis*.

> **To clarify:**
>
> Suppose in a selective program, 96 of 120 havo candidates are selected, and 30 of 50 mbo candidates are selected, there is an undesirable negative effect according to this rule.
>
> The selection of havo students is 80% (96/120); the selection of mbo students is 60% (60/100); the minimum for a fair ratio to mbo students is 64% (80% of 80%); 60% is lower than 64%, and therefore there is a negative effect according to the 4/5 rule.

To test the three criteria, five *fairness checks* are performed (see Figure 6), which are depicted in a *fairness plot* (see Figure 7):

1. **Accuracy equality ratio (ACC)** = (TP + TN)/(TP + TN + FP + FN) = accuracy: the ratio of correct predictions (both positive and negative) to all predictions - the number of students who correctly received a positive or negative BSA relative to the total number of students.
2. **Equal opportunity ratio (EO)** = TP/(TP + FN) = sensitivity: the ratio of correctly positive prediction to false negative prediction - the number of students who rightly received a positive BSA to the number of students who wrongly received a negative BSA.
3. **Predictive opportunity ratio (PO)** = FN/(FN + TN): the ratio of falsely negative predictions to all negative predictions - the number of students who falsely received a negative BSA to the total number of students who received a negative BSA.
4. **Predictive parity ratio (PPV)** = TP/(TP + FP) = precision: the ratio of correctly positive predictions to all positive predictions - the number of students who rightly received a positive BSA to the total number of students who received a positive BSA
5. **Statistical parity ratio (STP)** = (TP + FN)/(TP + TN + FP + FN): the ratio of correctly positive prediction and false negative prediction relative to all predictions - the number of students who cor-

[5] As far as we know, there is no European or Dutch equivalent of this criterion. Hence, in this study, we adhere to this rule.

rectly received a positive BSA or incorrectly received a negative BSA relative to the total number of students.

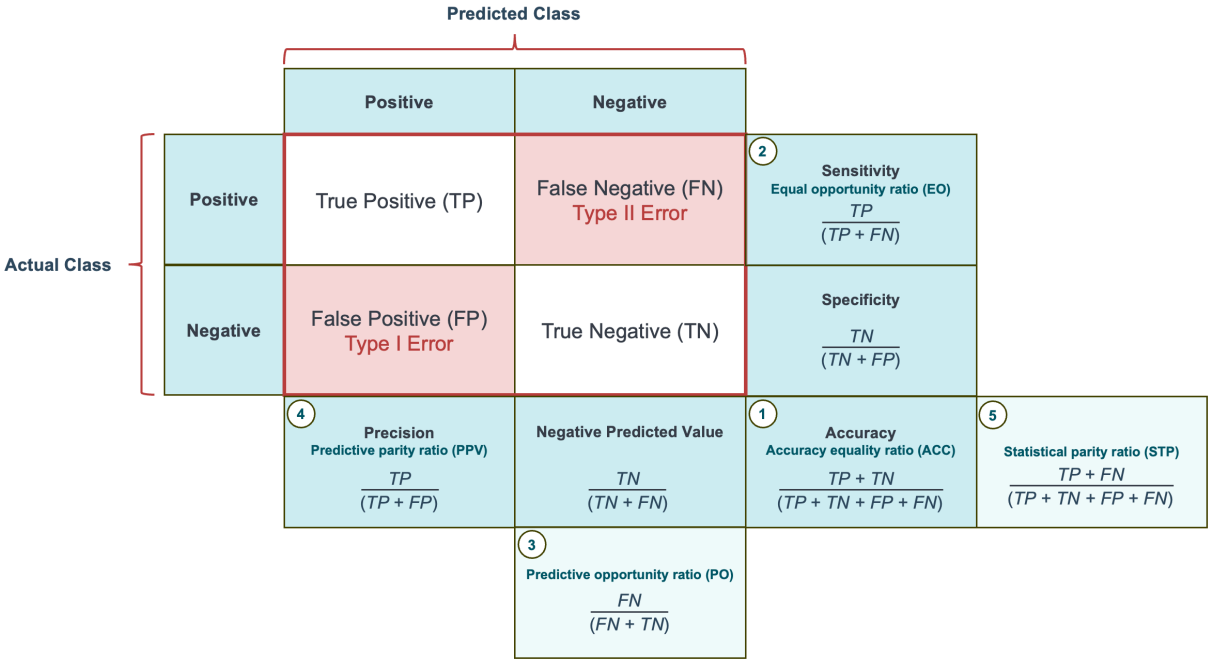

Figure 6: Confusion matrix with fairness checks

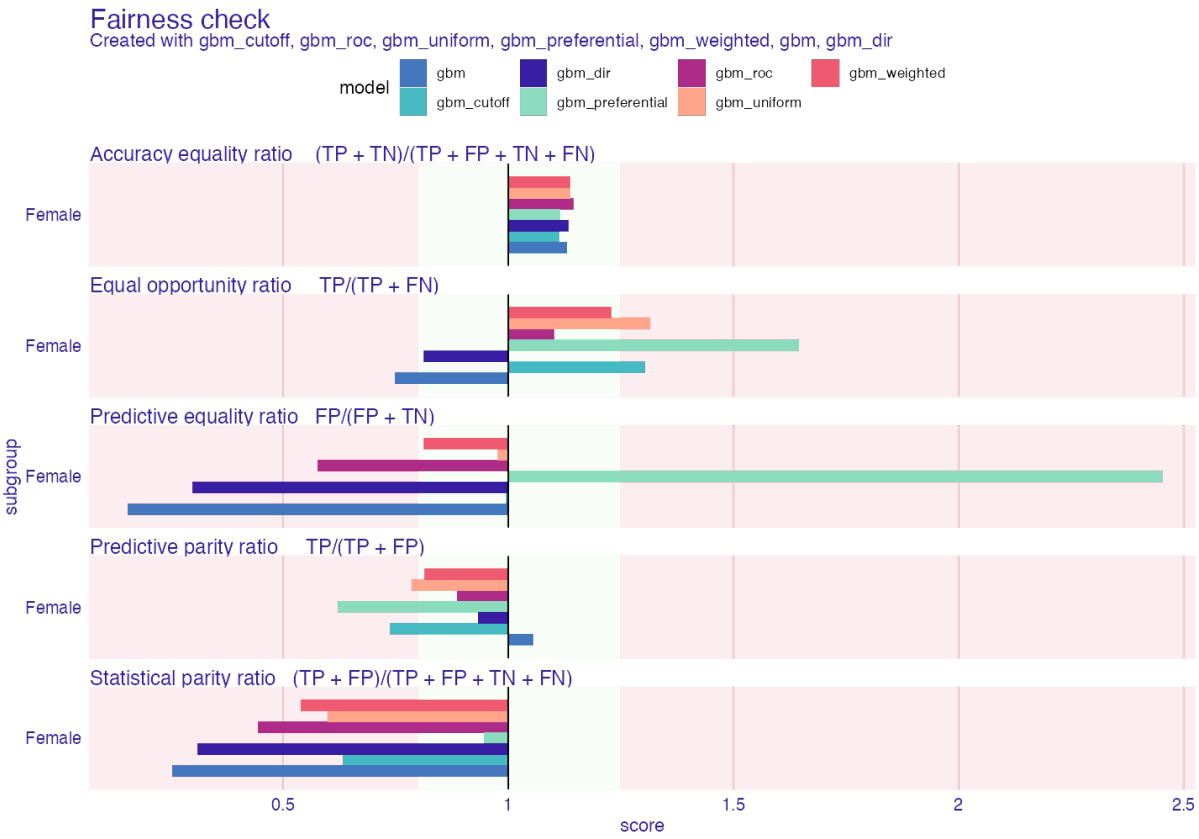

Figure 7: Visualization of fairness checks for the attribute gender (Source: fairmodels - dataset: German credit data)

Mitigation of bias

The models developed in the first phase with the `DALEX` package can be tested in the second phase for fairness to statistical minorities and - where necessary - corrected using the above ratios. For this purpose, there are methods a) prior to modeling (*pre-processing*): applying transformation of a distribution (*disparate impact remover*), weights (*reweighting*) or resampling of the data, or b) after completion of modeling (*post-processing*): reject option based classification pivot (*roc_pivot*) or adjusting cutoffs per group (*cutoff manipulation*). The outcomes are visualized using the `arenar` package (demo) and the `modelstudio` package (demo).

See Figure 8 for the complete research method.

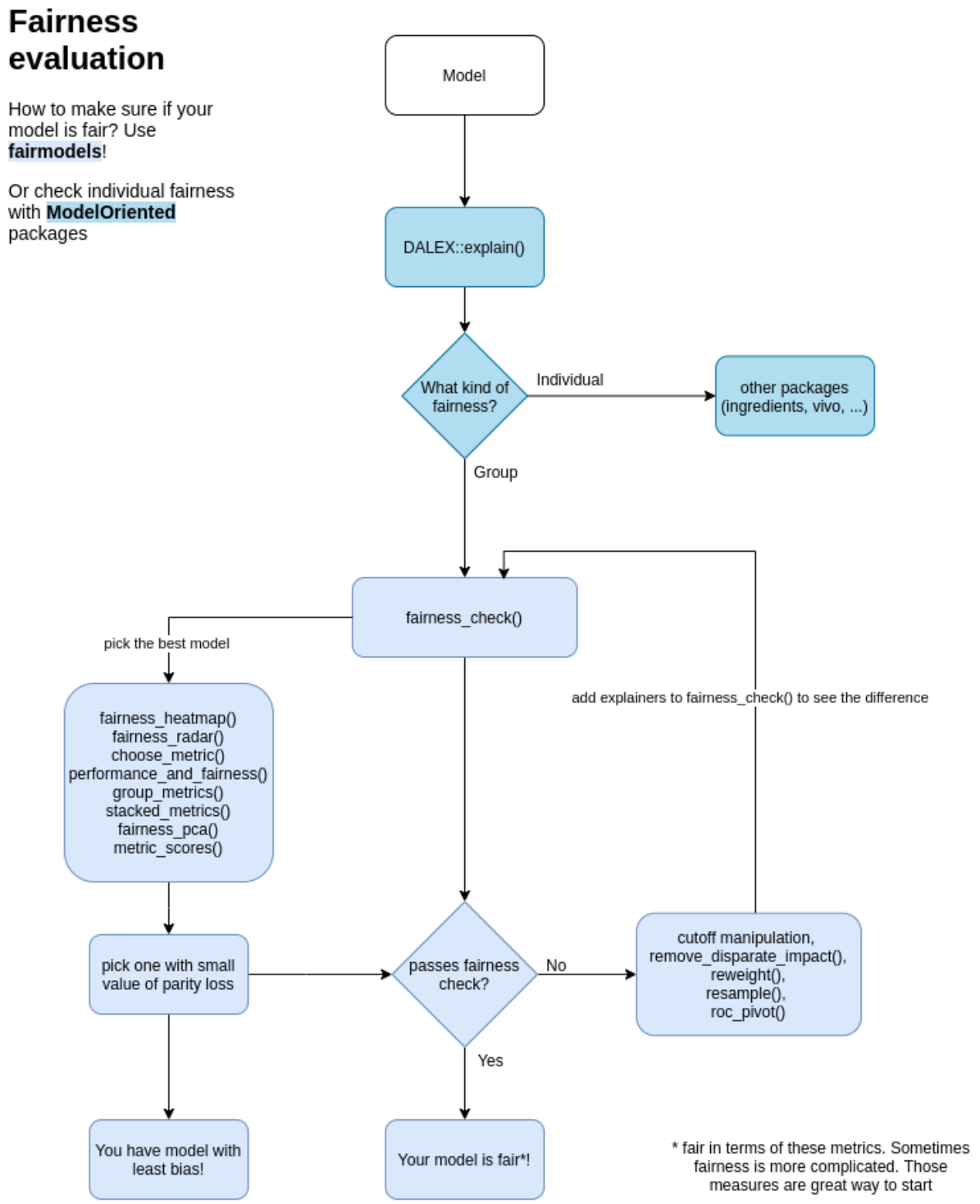

Figure 8: Group evaluation of fairness in modeling

6.3 Partial study III - Analysis of success in further study or in the labor market

This partial study answers the research questions from the third part of the operationalization of Equal opportunities for different talents to find a good place in society or equal opportunities for further study or position in the labor market.

For this, we will conduct a survey using CBS microdata. The details of this part of the research remain to be determined. In outline, the research involves using CBS to investigate the success of students who studied at THUAS and graduated with or without a degree:

1. In a **subsequent study program**: to which subsequent study did they go after leaving THUAS? At what level was this study? Did they earn a degree in this subsequent study? How long did it take them to obtain that degree?
2. On the **labor market**: After leaving THUAS, did they find paid work within three, eight, and 12 years? In what field of work did they find it? At what level of education and income?

In this study, we align with standard research methods of yet-to-be-explored labor market research, such as CBS and the Research Center for Education and the Labor Market (ROA).

A separate research plan will be developed for this study.

7 Expected results

The results of the study will be made available in the following ways:

- **Professional products**: Targeted advice to course management, research reports on subgroups, and methods for investigating equal opportunities in study data. Assurance of the application of insights through the Inclusion Office of THUAS, the implementation agenda of the Institutional Plan, and connection to the Equal Opportunities Alliance in the Hague region. Connection to national agendas through the SURF SIG Learning Analytics, Npuls' Study Data & AI Information Hub, SURF Study Data, and the Dutch AI Coalition for Education.
- **Inaugural address**: The research will form the basis for the lecturer's inaugural address (November 2024).
- **Publications**: "Equal opportunities in THUAS higher education" (possibly disaggregated by inflow, progression, and outflow); possibly derivative articles on sub-themes, such as "Influence of travel times on equal opportunities in higher education," "Study progress and success of students with Caribbean prior education" both in journals such as Higher Education and popular academic journals or magazines for target audiences such as Thema for higher education managers.
- **Presentations**: At conferences such as the SURF Education Days, the DAIR, the HO link, The THUAS AI Fest, etc.
- **Blogposts**: On the THUAS and SURF sites.
- **Source code**: The research source code will be made available via GitHub under CC license to conduct similar research: Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

8 Reproducibility

The research follows the FAIR principles for reproducibility of research results: **F**indability, **A**ccessability, **I**nteroperability, and **R**euse of digital assets. We relate this to source code, metadata, data, and professional products.

Table 7: Application of the FAIR principles

	Source code	Metadata	Data	Professional products
Findability	Will be published publicly on github	Will be published publicly on github	Access is described and published at github; sources that are not traceable to THUAS or students of THUAS are published	Articles will be made public; internal advisory reports will be made public or not, depending on the degree of traceability to THUAS.
Accessibility	Is freely accessible	Is freely accessible	Can be viewed upon request	Depends on the degree of public accessibility
Interoperability	Can be read using open software (R and RStudio)	Can be read using open software (R and RStudio)	Can be read using open software (R and RStudio)	Can be read with a PDF reader or via trade journals, presentations, blog posts
Reuse	Can be reused at will under CC BY-NC-SA 4.0 license	Can be reused at will under CC BY-NC-SA 4.0 license	Can be reused as far as public at will under CC BY-NC-SA 4.0 license	Can be reused as far as public at will under CC BY-NC-SA 4.0 license

References

Badou, M., & Day, M. (2021). *Kansengelijkheid in het onderwijs. Verkennend onderzoek naar factoren die samenhangen met onderwijs(on)gelijkheid* (p. 34). Gelijke kansen alliantie. https://www.gelijke-kansen.nl/documenten/publicaties/2021/10/05/verkennend-onderzoek-naar-factoren-die-samenhangen-met-kansengelijkheid

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness in Machine Learning Limitations and Opportunities*. fairmlbook.org. http://www.fairmlbook.org

Biecek, P., & Burzykowski, T. (2021). *Explanatory Model Analysis*. Chapman; Hall/CRC, New York. https://pbiecek.github.io/ema/

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T. P., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.

Code of Federal Regulations. Section 4d, uniform guidelines on employee selection procedures (1978). (1978). https://www.govinfo.gov/content/pkg/CFR-2011-title29-vol4/xml/CFR-2011-title29-vol4-part1607.xml

Copier, J. (2022). *Tussen idealen en dwalingen. Verhalen over onderwijs*. Garant.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2011). Fairness Through Awareness. *arXiv*. https://doi.org/10.48550/arxiv.1104.3913

Elffers, L. (2022). *Onderwijs maakt het verschil - kansengelijkheid in het Nederlandse onderwijs*. Walburg Pers B.V.

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: a new model for analysis of the educational process. *Educational Research*, *49*(4). https://doi.org/10.1080/00131880701717198

Fish, B., & Stark, L. (2022). It's Not Fairness, and It's Not Fair: The Failure of Distributional Equality and the Promise of Relational Equality in Complete-Information Hiring Games. *Equity and Access in Algorithms, Mechanisms, and Optimization*. https://doi.org/10.1145/3551624.3555296

Hardt, M., Price, E., & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. *arXiv*. https://doi.org/10.48550/arxiv.1610.02413

Huff, D. L. (1963). A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*, *39*(1), 81. https://doi.org/10.2307/3144521

Pedreshi, D., Ruggieri, S., & Turini, F. (2008). Discrimination-aware data mining. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/1401890.1401959

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-sts330

Wiśniewski, J., & Biecek, P. (2022). fairmodels: a Flexible Tool for Bias Detection, Visualization, and Mitigation in Binary Classification Models. *The R Journal*, *14*(1), 227–243. https://doi.org/10.32614/rj-2022-019

Version History

- 10-05-2023: version 0.9.0 - first draft version
- 18-05-2023: version 0.9.1 - second draft version; addition with sources, delivery and editing, methods of research, expected results
- 24-05-2023: version 0.9.1 - third draft version; extension of methods of research, ethical principles, reproducibility
- 25-05-2023: version 0.9.2 - tightening Introduction
- 27-05-2023: version 0.9.3 - addition with CBS microdata survey + additional illustration to clarify sub-studies
- 30-05-2023: version 0.9.4 - tightening terminology
- 31-05-2023: version 0.9.5 - addition of visualization tools in sub-study II
- 01-06-2023: version 0.9.6 - tightening the interpretation of equal opportunities based on Espinoza (2007)
- 02-06-2023: version 0.9.7 - supplement with appendix 1 - Data Science Ethics Checklist - version for the THUAS Ethics Advisory Committee
- 03-06-2023: version 0.9.8 - addition of background information to the DALEX package
- 06-06-2023: version 0.9.9 - update on version history errors
- 07-06-2023: version 0.9.9.1 - correction of typo's
- 07-06-2023: version 0.9.9.2 - additions on implementation of results
- 08-06-2023: version 0.9.9.3 - addition list of variables
- 30-06-2023: version 1.0 - incorporation of feedback from THUAS' Ethics Advisory Committee
- 05-07-2023: version 1.0.1 - incorporation of feedback from team IR&A, OKC
- 24-11-2023: version 1.0.2 - translation into English; new date for inaugural address
- 18-03-2023: version 1.0.2 - renaming the project
- 04-07-2024: version 1.0.3 - improvement of operationalisation

Repository

The source code for this document can be edited via GitHub. The professorship's knowledge base members can access and contribute content upon request.

Appendix 1 - Data science ethics checklist

8.1 A. Data collection

Focal point	Considerations
A.1 Informed consent	If there are human subjects, have they given informed consent, where the subjects themselves choose and have a clear understanding of the data use to which they are consenting? There are no human subjects in this study. The study data dataset is delivered anonymously to the lectorate by OKC. Elements removed by OKC are: student number, date of birth, first name and last name. Because of anonymization, the dataset does not fall under the GDPR because anonymous data are not personal data. Therefore, no consent applies to this dataset. The research's basis is the researcher's legitimate interest and THUAS.
A.2 Bias in data collection	Have we considered sources of bias that might be introduced during data collection and survey design and taken steps to mitigate them? See Data sources. The chosen sources come from the administrations of THUAS, public and administrative sources, and the national student survey. No additional sources were collected by the professorship based on any surveys of THUAS students. Gaps in source collection are those data that are informal: social contacts, motivation, expectations, etc. These data are not available or not accessed within THUAS (e.g., social interactions in an online learning environment). All administrations and surveys make a selection of information and are therefore biased. The degree of bias is the subject of research from the weighting of models and unexplained variance on possible outcome variables.
A.3 Limited disclosure of personally identifiable information	Have we considered ways to minimize the exposure of personally identifiable information (PII), such as through anonymization or by not collecting information irrelevant to analysis?

Focal point	Considerations
	The data is anonymized (see above). Information that becomes available by zip code is aggregated to neighborhoods and districts. To illustrate possible predictive power, fictitious student profiles will be created. The collection of specific data will be limited to those data known from literature or previous research by the lecturer.
A.4 Mitigation of downstream bias	Have we considered ways to test downstream results for biased outcomes (e.g., collecting data on protected group status such as ethnicity or gender)? These tests are the crucial subject of this study to determine the bias on these characteristics. For now, ethnicity is not included in favor of the SES-WOA characteristics.

8.2 B. Data storage

Focal point	Considerations
B.1 Data protection	Do we have a plan to protect and secure data (e.g. encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)? Data are stored in accordance with the college library guidelines of THUAS via SURF Research Drive. Any data transfer takes place using SURF Filesender, with proof of transactions stored separately. The code is stored separately in a private github environment for members of the Knowledge Circle of the Learning Technology & Analytics Professorship. Version control on the software (R) is done via the renv package in R.
B.2 The right to be forgotten	Do we have a mechanism by which individuals can request that their personal data be deleted? Because the data is anonymized, removing a student afterward is impossible because the student cannot be identified.
B.3 Data retention plan	Is there a schedule or plan to delete the data when it is no longer needed?

Focal point	Considerations
	The data will be archived in DARK store after the completion of the research and publication of the results after one year. THUAS does not yet have this facility, but the professorship will discuss this with the data stewards of the college library.

8.3 C. Analysis

Focal point	Considerations
C.1 Missing perspectives	Have we attempted to address blind spots in the analysis through collaboration with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)? The outcomes will be discussed with students and programs (teachers, program directors, educational advisors, etc.). For contact with students, we will partner with the Partner Up! program of the Global & Inclusive Learning Knowledge Center and the Inclusion Office of THUAS.
C.2 Dataset bias	Have we examined the data for possible sources of bias and taken steps to reduce or address these biases (e.g., perpetuation of stereotypes, confirmation bias, unbalanced classes, or omitted influencing variables)? This is the subject of this study: identifying bias and opportunities to reduce it for the benefit of students, teaching and educational policy at THUAS.
C.3 Fair representation	Are our visualizations, summary statistics and reports designed to fairly represent the underlying data? The visualizations, summary statistics, and reports aim to map bias, including the effects of adjustments correctly. This is an integral part of the software used: the `DALEX` and `fairmodels` packages.
C.4 Privacy in analysis	Have we ensured that data with PII is not used or displayed unless necessary for analysis? Yes. See response to A.3.
C.5 Verifiability	Is the process for generating the analysis well documented and reproducible if we discover problems in the future?

Focal point	Considerations
	Yes. Please refer to the section on Reproducibility

8.4 D. Modeling

Focal point	Considerations
D.1 Proxy discrimination	Have we ensured that the model does not rely on variables or approximations for variables that are unfairly discriminatory? This is the subject of this study.
D.2 Fairness between groups	Have we tested the model results for fairness with respect to different affected groups (e.g., tested for unequal error rates)? This is the subject of this study.
D.3 Feature selection	Have we considered the effects of optimizing for our defined variables and have we considered additional variables? The optimization of the variables and their effect on possible prediction is the subject of research (via ceteris paribus analyses and bias analyses). Possible additional variables were partially considered and discarded due to their lack. In addition, additional ideas about variables may arise as the research progresses, which will be reported.
D.4 Explainability	Can we explain in understandable terms a decision made by the model in cases where justification is needed? Yes. The `DALEX` and `fairmodels` packages aim to precisely break down the internal workings of analyses and models in favor of "glass box" models.
D.5 Communication about bias	Have we communicated the model's shortcomings, limitations and biases to relevant stakeholders in a way that can be widely understood? The limitations of the study will be included in the reports/articles and differences forms in which we will disclose the study results. See Expected results

8.5 E. Deployment

Focal point	Considerations
E.1 Monitoring and evaluation	How do we plan to monitor the model and its effects after it is implemented (e.g., performance monitoring, regular checking of sample predictions, human assessment of high stakes decisions, assessment of downstream effects of errors or low reliability decisions, testing for concept drift)? The study is expected to provide insight into variables on which bias exists in THUAS. These may be adopted in teaching and educational policy, depending on the results. This may become part of the program to implement the Institutional Plan. We will discuss this with the CvB and the Director of Strategy.
E.2 Repairing and preventing damage	Have we discussed with our organization a plan of action in case users are harmed by the results (e.g., how does the data science team evaluate these cases and update analyses and models to prevent harm in the future)? In the lecturer's experience, good communication of outcomes is essential, especially with the press. In presenting and discussing the results, multiple perspectives on fairness will be addressed to avoid becoming dominant. In addition, we are monitoring the use of these insights. Our influence on this - as with any scientific research - is limited.
E.3 Rollback	Is there any way to disable or reverse the model in production if necessary? The study does not develop models that will go into production. This may be part of a follow-up project and will be addressed then.
E.4 Inadvertent use	Have we taken steps to identify and prevent unintended use and misuse of the model, and do we have a plan to monitor this once the model is in place? See the answer to question E.2

Source: Data Science Ethics Checklist, created with deon.