Beyond Accuracy: Rethinking Outlier Detection in Asylum Data

KRISTIN KALTENHÄUSER, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
TIJS SLAATS, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
MICHAEL MULLER, IBM Research, Cambridge, United States
NAJA HOLTEN MØLLER, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

In asylum decision-making, algorithmic tools are developed to optimize decision-making processes, for example, to streamline the application process. Accuracy, defined as correctly categorizing cases against a ground truth dataset, is a main concern. Data scientists employ outlier detection algorithms (ODA) to boost accuracy by fine-tuning datasets and discarding cases deviating from mathematical norms to prevent over-fitting. However, algorithms increasingly face scrutiny for possibly harmful effects on marginalized communities. Just as supervised machine learning is understood as a method to produce decision boundaries, ODAs are understood as methods to produce outliers – who sometimes are human beings. Organizations that use ODAs put people at risk of becoming marginalized by discarding them from a dataset, which motivates this paper. We critically engage with outlier handling as a significant juncture that molds representation within datasets. Using a dataset of asylum case summaries, we investigate the when individuals are constructed as outliers, by comparing three different ODAs and four different parameters for data modeling. This is significant because inclusion in datasets used for training automated decision tools and as ground truth, like the one under study, is crucial for ensuring transparent decisions. This study finds little overlap between the outliers constructed by the three algorithms and different data representations. This means that the outlier designation heavily depends on the choice of ODA and the parameters used in modeling the data; yet, data analysts rarely give much thought to this choice. Instead, decisions are primarily guided by technical considerations such as accuracy. Algorithms shape a specific understanding of key concepts such as text topics and norms, urging us to reconsider the design of algorithmic tools beyond merely achieving high accuracy and focusing on what makes an algorithmic decision tool useful for the particular context and the consequences for those who are constructed as outliers.

CCS Concepts: • **Applied computing** → **Computing in government**; *Document management and text processing*; • **Theory of computation** → **Design and analysis of algorithms**; • **Computing methodologies** → **Natural language processing**; Machine learning algorithms; • **Human-centered computing**;

Authors' Contact Information: Kristin Kaltenhäuser, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: krka@di.ku.dk; Tijs Slaats, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: slaats@di.ku.dk; Michael Muller, IBM Research, Cambridge, Massachusetts, United States; e-mail: michael_muller@us.ibm.com; Naja Holten Møller, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: naja@di.ku.dk.

ACM 2832-0565/2025/04-ART5
https://doi.org/10.1145/3718986

Additional Key Words and Phrases: Outliers, asylum, outlier detection, AI, algorithmic decision-making, public decision-making, LOF

ACM Reference Format:
Kristin Kaltenhäuser, Tijs Slaats, Michael Muller, and Naja Holten Møller. 2025. Beyond Accuracy: Rethinking Outlier Detection in Asylum Data. *ACM J. Responsib. Comput.* 2, 1, Article 5 (April 2025), 28 pages. https://doi.org/10.1145/3718986

1 Introduction

Critics warn of the disparate impact of **Artificial Intelligence (AI)** on marginalized communities and the risks of excluding these perspectives in the training data and development of AI systems [104], in domains such as language models [18], image recognition [26], public decision-making [8, 80], and others [19]. Despite a history of work addressing issues of inclusion and exclusion [3, 36, 37, 84], only a few HCI studies have focused on identifying the individuals at risk of being marginalized by such AI systems and understanding the practical mechanisms and conditions of marginalization [49, 62, 74, 102, 107]. Lessons from these considerations need to be considered when introducing decision-support to high-stakes domains such as asylum, the context in which we operate. Governments worldwide experiment with algorithmic tools and AI as an attempt to reduce overwhelming costs of handling large case-loads [17, 50], and in the hopes to reduce or remove human judgment bias and make administrative decision-making more just and optimized [43, 48, 51]. Efforts to employ AI for predicting asylum outcomes have emerged in various countries, including the US, Canada, and Australia [28, 31]. Researchers such as Louise Amoore, emphasize that AI inherently mirrors historical data, thus perpetuating past patterns in decision-making [7]. Consequently, AI tools in asylum may inadvertently repeat and amplify potentially biased past decisions. Moreover, if the parties who design the automated tools lack sufficient training data, they put marginalized communities at risk for opaque decisions and incomprehensible reasoning.

To gain a better understanding of the development process, the adverse effects of automated decision tools, and the potential targets of these effects, **human-computer interaction (HCI)**, **computer-supported cooperative work (CSCW)**, and critical data studies communities have conducted numerous and significant ethnographic studies. These studies delve into how humans engage with data, algorithms, and algorithmic outcomes when these AI models are developed [27, 73, 74, 87, 99, 123]. These studies emphasize the importance of interrogating data science practices [70] through ethnographic studies, as a critical aspect of understanding how datasets are shaped by algorithms and the individual [71]. We seek to contribute to this strand of research by employing quantitative experiments with different algorithms and algorithmic parameters, investigating how the data cleaning process, and specifically outlier handling, shapes representation in datasets and possibly contributes to the marginalization of individuals.

Outliers are individual or small groups of cases that diverge significantly from the *norm*. While most analyses of outliers focus on detection and amelioration of outliers [30, 63] - including criminalization of certain classes of outliers [81, 122] - in this paper we invert this strategy to problematize the norm(s) against which outliers are compared. As such, outliers raise the question of what the implied *norm* constitutes. **Outlier detection algorithms (ODA)** employed to distinguish between normal and outlier data points impose a structure onto the data. This structure, termed a data model, abstracts and simplifies the data. This yields the mathematical concept of a norm, which identifies a subset of data points as outliers. Once labeled as outlier, these points are often excluded from the dataset. Therefore, outlier detection essentially determines which individuals are represented in a dataset and which are not.

Data cleaning is a pivotal phase in constructing any AI system. Many data scientists acknowledge that a substantial portion of their professional efforts revolves around this phase [96], despite it being commonly perceived as uninteresting [58]. For data scientists, clean data signifies "data that is free from errors, outliers and other inconsistencies" [114, p. 262]. This suggests that decisions made during this stage dictate the inclusion or exclusion of data points in a dataset, ultimately determining whose experiences are represented in subsequent analyses or algorithmic systems [39]. *However, existing data science literature indicates that while considerable effort is dedicated to identifying outliers, information is scarce regarding the criteria for deciding whether to retain or remove these outliers from the dataset.* Accuracy, defined by correctly categorizing cases against a ground truth dataset, is a main concern. Data scientists employ outlier detection algorithms (ODA) in the data cleaning process to boost the accuracy of their data model by fine-tuning datasets, and discarding cases deviating from mathematical norms. However, the decision of which data points to exclude from the dataset holds critical importance on various fronts. Ensuring comprehensive representation in a dataset, particularly when each data point represents a human experience (as opposed to, for example, temperature measurements over time) is essential for averting the marginalization of communities and individuals when the dataset is used as training data for decision-making algorithms [57]. The manifestation of this marginalization varies from one case to another. The issue lies in the subtle decisions made by developers of AI tools, which ultimately shape the final decision prediction. For instance, if the tool lacks sufficient data to make a confident decision, will it communicate uncertainty or default to a predetermined outcome? Marginalization therefore extends beyond conventional factors like race and gender, as demonstrated in Kaltenhäuser et al. [57]. It encompasses potentially unfair treatment of communities and individuals sharing other common characteristics, such as families of diverse nationalities in our context.

In line with this perspective, scholars such as Rieder [98] posit that using automated decision tools "don't take away human discretion, but shift it from a definition of outcomes to the definition of procedures, mechanisms, and techniques that produce outcomes" (p.113; see also [6, 101]). That means that the choices that data analysts make in the data cleaning process to optimize procedural parameters, such as accuracy, significantly influence the decisions produced by automated decision tools. It is therefore imperative to closely examine how these procedural choices impact the ultimate decision outcome.

In this paper, we focus on the domain of asylum decision-making to illustrate how human experiences come to be included or excluded from a dataset as outliers in the data cleaning process. Understanding the conditions through which data points are labeled as outliers and potentially excluded from a dataset that might serve as a basis for future decision-making can help mitigate inequitable outcomes. This is particularly crucial in domains such as asylum decision-making, where the target population is inherently vulnerable from the outset. Excluding certain individual experiences as outliers could lead asylum practitioners and policymakers to overlook important patterns and trends that could inform more equitable policies, interventions, or decision practices. We break down and conceptualize the computational production of outliers as part of the data cleaning process to gain an understanding of the conditions that shape the possible exclusion of lived experiences.

Using a dataset comprising asylum decision summaries from Denmark, we demonstrate the impact of various data representation parameters and ODAs on the classification of data points as outliers. Additionally, we critically analyze the different algorithms we employ for outlier labeling and examine the underlying calculations in relation to the case files and the concepts the algorithm enacts. We ask: What are the conditions of the production of outliers and hitherto representation in datasets, and how can we accommodate a human-centered perspective onto outlier detection?

To address this question, we conduct quantitative experiments that involve testing different parameters in what we coin *data featuring practice*, as well as the choice of ODAs. Through this process, we showcase how different approaches to featuring and outlier detection result in significant differences in determining which individuals are excluded from a dataset and subsequent analyses as outliers.

2 Related Work

2.1 The Work of Data Science

Recently, attention in data science has begun to shift toward human-centered questions, methods, and analyses. Shneiderman [105] developed an influential account of human-centered AI. The related field of **human-centered data science (HCDS)** examines [9] and interrogates [72, 113] the roles of humans in creating data science applications, and the needs of humans who use those applications, or who are indirectly impacted by those systems (e.g., [44]). The goal is to complement more technological approaches, moving from "how it works" toward "how and why it was made" [46, 70, 74, 99] and "how it affects humans, their work, their families, their communities, and their societies" [12, 60, 100, 101, 106], and "how can it be improved to meet human needs" [9, 20, 41, 59, 70].

There are numerous significant ethnographic studies on how data science workers engage with data, algorithms, and algorithmic outcomes [73, 74, 76, 87, 99, 123]. Möller et al. [78] illustrate that the task of producing quality data often falls to individuals other than data scientists, such as clerical workers and caseworkers. They argue that understanding the limitations of datasets significantly depends on data scientists' awareness of the work practices involved in the initial data production. Passi and Jackson [87] joined a data science team for six months to investigate how trust in data science project outcomes is negotiated in an organizational context. Building on the work of Feinberg [46], Muller et al. [73] interviewed data science practitioners to show the different ways they engage with data, such as, e.g., designing or even creating data instead of simply *discovering* it. While Sambasivan et al. [99] use the concept of *data cascades* to characterize the effects of human intervention in the data science process, Muller and Strohmayer [74] approach it through the notion of *forgetting practices* and *data silences*. Miceli et al. [66] delved into the work of data annotation, highlighting it as a sense- and meaning-making activity influenced by power relations, and demonstrating how data is shaped by power dynamics. The workshop organized by Pine et al. [90] conducted a cross-domain analysis on the demanding nature of data work, which often necessitates skilled and intensive efforts. However, it is frequently overlooked when analyzing the development of AI tools [83].

Turning to the modeling and algorithmic work of data science workers, Passi and Barocas [86] shed light on the negotiation inherent in translating real-world problems into data science problems. Following a similar strategy of analysis, Pine and Liboiron [91] showed that the definition of *the data* to be analyzed may also involve negotiations and questions of the power of multiple stakeholders.

Paullada et al. [88] surveyed studies about dataset design and development, identifying common issues such as representation, which involved biases, in terms of inclusion and exclusion of individuals. Mentis et al. [65] and Feinberg [46] showed that representations are often *tuned* and *crafted* to be fit-for-purpose for later stages in the analysis. Representation challenges appear in the data collection phase, but also manifest significantly during the data cleaning process, specifically the outlier handling phase – a pivotal moment in determining representation, we show.

These studies emphasize the significance of human discretion in shaping the landscape of data science problems, influencing data representation, and impacting (the interpretation of) outcomes.

While existing studies highlight these aspects, they often lack a detailed exploration of the precise mechanisms through which human choices exert influence, particularly on individual data points. We add to this body of research by delving into practices of outlier handling in datasets and ensuing ML systems. Our approach adds to these ethnographic studies, employing quantitative experiments involving various parameters and outlier detection algorithms. We aim to demonstrate how different choices in outlier handling lead to substantial variations in the exclusion of specific data points during subsequent analyses, taking a human-centered approach to data science.

2.2 Mitigating Algorithmic Harm through the Analytical Lens of Outliers

The concept of an *algorithm* as an analytical object has been a subject of prolonged debate regarding what aspects should be scrutinized when investigating its societal impact. In a recent commentary, Suchman [109] cautioned against uncritically treating AI as a monolithic entity. Instead, she advocated for a nuanced perspective, considering AI as a composite of various techniques and processing steps. This stance aligns with Dourish's earlier exploration of algorithms [40], emphasizing their interrelation with other computational elements like data and code.

Critical scholars have proposed various frameworks to investigate algorithmic harm in the complex ML lifecycle, such as [110]. Suresh and Guttag [110] map the ML lifecycle to potential sources of harm. In the data preparation phase, a potential source of harm is representation bias, when the data underrepresents some communities (e.g., [107]). This paper adopts this framework by narrowing its analytical focus to the data preparation phase. Rather than examining potential biases in an ML system as a whole, our analysis targets a succinct step: data cleaning and outlier handling. This approach allows us to delve into the intricate details of this particular phase, shedding light on its implications by experimenting with various parameters and contributing to a more granular understanding of algorithmic processes.

Researchers across various domains have explored the role of outliers within their respective fields, demonstrating how individual cases that deviate from established norms are frequently overlooked by data models relying on extensive datasets. In the realm of education, Treviranus [116] advocates for heightened attention to outliers within an increasingly data-driven U.S. education system to accommodate the diverse needs of learners. She asserts, "Design based on metrics for the norm may be detrimental to the margins, but the converse is not true." Similarly, Deitz [38] investigates outliers in city planning and identifies an outlier bias against people with disabilities. The study reveals that individuals not conforming to societal or algorithmic norms, such as those with disabilities, are systematically excluded. Further, a psychological study by Pollet and van der Meij [92] explores the effects of retaining or discarding outliers known not to be noise in hormonal data, finding significant differences in statistical conclusions based on these choices. Bardzell recommended to begin analysis *from the margins*, both to strengthen the epistemological basis of the research and to reduce the likelihood of excluding outliers [13]. Addressing the issue of inclusive digital infrastructures, Feinberg et al. [47] explore alternative designs for database infrastructures, embracing residuals as data points lying outside conventional values, and centering rather than marginalizing them.

These studies underscore the critical importance of considering outliers, especially in the context of large datasets. They highlight the impact that a normative perspective can have on outliers, such as street layouts designed for able-bodied individuals and without accounting for the diverse range of physical abilities [113]. Building on this premise, we assert that the production of outliers often stems from choices made by data scientists during the data cleaning phase.

In addition to the aforementioned studies, critical scholars have drawn attention to the significance of data points that defy predetermined categories or data models derived from large datasets. Thylstrup et al. [115, Chapter on "Outliers"] illustrate this concept using the example of

gender datafication, highlighting the marginalization of non-binary individuals in datasets where gender is encoded as binary (see also [34, 107]). The authors emphasize that representation not only mirrors reality but actively shapes it. The establishment of an inclusive digital infrastructure therefore becomes crucial for fostering a more inclusive world, necessitating an acknowledgment of the complete spectrum of values within a category.

Furthermore, there is a growing chorus among scholars cautioning against the assumption that more data necessarily equates to greater truth value. boyd and Crawford [24] argue that tracing individual cases can offer insights that might be overlooked in the analysis of large datasets. In a similar vein, Welles [120] advocates for the inclusion of minorities in Big Data analysis not as deviations from statistical norms, but as reference categories. Likewise, Strohmayer and Muller [108] contend that data scientists have yet to thoroughly explore the margins and outliers within their datasets.

Bechmann [15] analyzes the political dimension of this discourse, asserting that when processing data as humans, attention must be paid to representation, ensuring that all individuals contribute to data processing on equal terms. This becomes compromised when outliers are filtered out or data is adjusted for abnormality.

These investigations demonstrate the repercussions of outlier or residual labels within a dataset or digital infrastructure, illustrating how such positioning can result in marginalization. Expanding upon this foundation, our study delves into the circumstances surrounding the algorithmic labeling of outliers and explores the impact of seemingly minor decisions made by data scientists in shaping this categorization.

2.3 Navigating Outlier Ambiguity in Data Science Research

The strand of data science research that deals with *identifying* and *handling* outliers – i.e., *outliering* [5, 77] – emphasizes the importance of *clean* data as a prerequisite of a *good* data model with a high accuracy [32]. Accuracy, used in this study as the ability to correctly categorize cases against a ground truth dataset, is among the metrics used to evaluate data models [114]. Data scientists employ outlier detection algorithms (ODA) to boost accuracy by fine-tuning datasets. Azeroual [10] writes about how outliers can "distort the results of the analysis. These must be cleaned to get the best results" (p.4). However, the data analyst often faces uncertainty when deciding how to *handle* these data points. Outliers can be caused by various factors, such as measurement errors or data entry mistakes, in which case they are removed from the data [32] or replaced with a *reasonable* value [111]. Outliers can also represent a genuine variation in the data and therefore an important data point.

In data science literature, there is an inconsistency in the use of terms like *outlier, anomaly,* and *noise*. For instance, Azeroual [10] refers to outliers as "zero values, duplicates, outdated, invalid, and unformatted data" (p.4). However, in the same paper, the author mentions that "outliers in a graph can often be recognized as conspicuous extremes" (p.9). Consequently, when encountering discussions on outliers in computer science literature, the term may encompass various interpretations, ranging from extreme values lying outside a specific range of standard deviations from the mean, to missing values or mere measurement errors. Moreover, contemporary machine learning algorithms designed to detect outliers adopt a more nuanced definition, considering outliers as data points deviating from data models beyond the mean. These models may encompass factors such as points not belonging to any data point clusters or those failing to conform to the established patterns of other data points. Outlier detection algorithms (ODA), encompassing both **machine learning (ML)** models and traditional algorithms, are widely utilized to identify outlier points within datasets and are a focus of the following analysis, in which we select three different ODAs and compare their results.

Data science literature often falls short in providing guidance on how to distinguish whether to treat a data point as noise or retain it. Despite underscoring the significance of outlier handling, resources like the widely used book "Pattern Recognition" [114], a fundamental reference for both students and practitioners delving into ML's mathematical and algorithmic aspects, allocate a section to *outlier removal*. This heading itself implies a predisposition towards discarding outliers, reinforcing a perception that outliers reflect data that is somehow deficient or damaging. The text continues along these lines, stating, "Points with values very different from the mean value produce large errors during training and may have disastrous effects. These effects are even worse when the outliers are the result of noisy measurements. If the number of outliers is very small, they are usually discarded" (p. 262). While acknowledging that outliers may result from causes beyond noisy measurements, the passage neglects to further explore these alternative causes. It underscores the necessity of removing such data points, even when they may be valid, without providing a comprehensive examination of the circumstances under which outliers could be retained.

Data science literature extensively covers outlier identification, yet offers scarce information on the critical aspect of determining *how* to handle these outliers. Campesato [29] succinctly notes: "Although the decision to keep or drop outliers is your decision to make, there are some techniques available that help you detect outliers in a dataset." [p.39]. The magisterial book "Outlier Analysis" [4] spans 500 pages, delving into outlier identification across various data types, including text, image, and time series data. However, the book overlooks a thorough exploration of strategies for dealing with outliers, leaving a notable gap in the discourse on this crucial aspect of data analysis.

There has been significant progress by the date science community in enhancing existing outlier detection techniques and creating taxonomies to help data science practitioners navigate this rapidly developing field [21, 85, 118]. Nevertheless, there is a dearth of research on the crucial aspect of determining whether outliers should be considered as noise or valid data points. An illustration of such a rare study is Kandel et al. [58]. They describe their process of iteratively removing points that visually appear as outliers. They use visualization to assess data quality and *detect* outliers in a way that assigns the human a more prominent role than if using ODAs: "Some outliers seem to be valid data (e.g., data from the District of Columbia are very different from data from every other state). Others seem suspicious (criminals may vary in age from teenagers to older adults, but apparently babies are also committing crimes in certain states). John iteratively removes those outliers he believes to be dirty data (e.g., criminals under 7 and over 120 years old)."

The scarcity of literature in data science addressing decision-making in outlier handling, coupled with the perceived negative impact of outliers on data model accuracy and the inconsistent use of related terms, places the responsibility on data science practitioners to determine the appropriate course of action for individual data points. While practitioners have an extensive toolbox for outlier identification, there is a notable lack of guidance on how to handle them and the repercussions of such decisions. Existing studies tend to focus more on the identification of outliers rather than their significance for the analysis or the studied phenomenon, or how different techniques may yield varying outlier outcomes. This study aims to bridge this gap by exploring the significance of outliers and how different methodologies produce distinct outlier results.

3 Method

The subsequent section first provides an insight into the origins and characteristics of our case study and the dataset under scrutiny, our take on how we can apply human-centered data science in the asylum domain. Next, we describe the data processing and transformation steps. Table 1 illustrates a summary of this. The goal of this section is to provide a detailed account of the underlying assumptions and calculations of the applied data processing steps, which will inform

Table 1. Data Processing and Analysis Steps

Phase	Steps	Result
0. Data Preparation	Removing unintelligible files, tokenization, stemming, etc.	17,263 asylum cases as lists of word tokens
1. Data Feature Engineering	a. TF-IDF vectorization	17,263 TF-IDF vectors
	b. Case transformation into topics (dimensionality reduction) using LSA	4 datasets transformed into d=2,4,8,30 dimensions (selected using topic coherence scores)
2. Outlier Detection	a. Applying 3 ODAs to the 4 datasets resulting from phase 1 b. Analyze intersection of the 12 outlier datasets from 2.b c. Inspecting 5 randomly selected cases labeled as outliers	list of outliers for all 3 ODAs for each of the 4 datasets

the subsequent section, laying out the findings. The following sections are arranged around two phases, as can also be seen in Table 1. The table additionally shows a phase 0, which involves some necessary text pre-processing steps. We provide a more detailed description of these steps, as well as the code used, on GitHub:

(1) Phase 1: Data Featuring - In Section 3.2, we elaborate on the conversion of textual data into numerical vectors, a crucial initial stage in the translation of real-world information into mathematical calculations by algorithms.
(2) Phase 2: Applying Outlier Detection Algorithms - In Section 3.3, we delve into the various Outlier Detection Algorithms (ODAs) applied to our transformed data, shedding light on the process of outlier production and how they hinge on the concept of a norm embedded in the ODA.

3.1 Our Case Study – a Dataset of Danish Asylum Case Summaries

The **Danish Refugee Council (DRC)**, an NGO with counseling responsibilities for refugees and veto rights in the Danish asylum decision process, provided our research project with a collection of approximately 30,000 documents, including Word files, scanned PDFs, temporary, and meta-data files, containing information about the asylum cases of the Danish **Refugee Appeals Board (RAB)** in the years ranging from 1996 – 2020. To facilitate in-depth analysis, our research project established a legal partnership with the DRC. Of the case files, the scanned PDF files were not machine-readable and we had to rely on **object character recognition (OCR)** to convert them into text files. After removing unintelligible and corrupted files, we ended up with around 17,700 text files accessible for further data analysis. A majority of the documents contain several types of information:

— demographic data about the applicant
— information about the procedural history of the case in the Danish asylum system
— a summary of the account of the asylum seeker about their family situation and reason for migrating from their country of origin
— interview minutes from the DIS and RAB
— decision and legal reasoning, including citations of legal texts

Not all documents contain all the information listed. This discrepancy can be attributed, in part, to recent organizational changes in the collaboration between the various Danish decision entities.

For instance, a different caseworker system was introduced to facilitate information exchange between these organizations.

To understand the nature of the dataset, we will provide a brief description of the context and origin of our data in the Danish asylum system. The asylum system in Denmark is two-tiered. When an asylum seeker arrives in Denmark, the first data about them is collected by the Danish police. A caseworker of the **Danish Immigration Service (DIS)** then makes a decision on whether asylum is granted or not, after conducting one or more interviews and other data collection. Rejected cases are automatically appealed to the RAB, which makes it the second decision instance of the Danish asylum process. A large portion of the decisions go to the appeals board (RAB), which serves as a quasi-judicial body in which decisions are made by a group of three people.

Governments worldwide, including Denmark, are increasingly interested in leveraging AI for decision-making processes [28, 31, 43] - in part as response to ever-increasing caseloads [17, 50, 54, 93], and in part in an effort to reduce human biases in decision outcomes [48, 51].[1] The Danish government's 2019 National Strategy for Artificial Intelligence specifically highlights the use of AI in supporting case workers in the social sector. The strategy outlines that "Case workers can get decision support based on knowledge from previous cases. [...] With artificial intelligence, data can be used in new ways to improve the quality of service, make task solving more efficient and increase general trust in the public administration" [2, p.55f]. Rather than replacing case workers, the strategy emphasizes AI as a tool to enhance efficiency and service quality. However, it does not address the potential tensions that may arise from this integration and these possibly contradictory goals. Currently, in Denmark's asylum sector, data and algorithms are employed to assess the credibility of asylum seekers, playing a significant role in the decision-making process. Nielsen et al. [82] have examined the implementation of AI in evidence gathering, specifically how mobile phone data of applicants is analyzed. The algorithm used for this task is an off-the-shelf product supplied by a Swedish company. In summary, while AI is increasingly being integrated into decision-making processes with the goal of improving efficiency and reducing human bias, particularly in sectors like asylum and social services, these initiatives raise unresolved concerns about the potential conflicts and challenges that arise from balancing efficiency, fairness, and the human role in decision-making.

In the following, we provide an overview of the two phases of data processing we employed.

3.2 Phase 1: Data Featuring - Turning Text into Numbers

A critical premise of the research project we are part of, is the commitment to understand the inner workings of data science techniques and *map* these back onto reality. Taking seriously that data science is a work practice, we look at popular strategies to translate text into numbers as data *featuring*. ML makes things such as asylum decision summaries calculable by employing various strategies to turn their text into numbers. It is crucial to understand what factors are incorporated into these calculations. Accordingly, this paper aims to examine the common strategies used to achieve this. To guide our choice of numerical text representation, we critically examined two approaches previously used in our project on asylum decision-making:

(1) Hamilton Byrne et al. [55] and Piccolo et al. [89] treated each case as an object with extractable properties such as gender, country of origin, entry date, and applicant age. These quantified properties formed a data point, limited by the analyst's chosen taxonomy and categories.

[1]However, Rieder [98] noted that an algorithmic approach does not remove bias, but simply changes the point in the case protocol where bias may impact - i.e., from human case-worker to human creator of the algorithm (see also [74]).

(2) Piccolo et al. [89] viewed text documents as lists of word frequencies, a common method for representing documents as numerical vectors. This approach allows for discovering unforeseen text features and categories, and facilitating techniques like topic modeling.

We also considered deep neural networks, such as those used in modern **large language models (LLMs)** like GPT and BERT. Each method for making text calculable has its own advantages and limitations. To thoroughly explore and question the inner workings of algorithms, we made strategic choices based on the method's popularity, the desire to minimize assumptions made by data analysts, and the feasibility of providing detailed insights into algorithmic processes. Feature extraction – which is a kind of featuring – where specific aspects of the text are identified and used for representation, is a common method. However, it is constrained by the data analyst's assumptions about which features are truly representative. To minimize assumptions about representationality, we opted for the second approach, which relies on word frequencies and semantic similarity to represent and compare cases. This method enables a deeper exploration of the text's potential, allowing algorithms to uncover criteria that may not be immediately apparent to humans. Additionally, semantic similarity is a key technique in computational text processing in general. Although LLMs are built on complex mechanisms like transformers and attention mechanisms, which capture deeper contextual relationships between words than word frequencies and semantic similarity, we excluded this approach from our study due to its complexity. While LLMs go beyond basic semantic similarity, they still focus and operate on the form of the text rather than its meaning, highlighting one of the key points we aim to address in this paper.

The subsequent section will delve into the rationale behind turning text into numbers using word frequencies.

3.2.1 Counting Word Frequencies in Documents: ***Term Frequency - Inverse Document Frequency (TF-IDF)****.* The TF-IDF algorithm transforms text documents into word frequency lists by highlighting words that occur frequently within a document, yet rarely across the entire collection. This method therefore aims to identify terms that represent the content of a document collection (corpus). This data representation encodes lexical information, meaning the words that the caseworker uses to describe the information about an asylum seeker and their journey through the Danish asylum system. For example, in one of the cases the applicant provided information about his uncle threatening to kill him because of a conflict about land ownership. The form in which this information is represented computationally is by a list of the following words:

> [conflict, uncle, land, inheritance, father, death, uncle, kill, applicant] (translated from Danish).

This example shows how semantic and syntactic properties, such as those dependent on word order, are lost in this transformation. Despite these disadvantages, TF-IDF is still among the most popular techniques to embed texts into numerical vectors across different domains. TF-IDF remains widely adopted for text-to-numerical vector conversion. A literature review found over 83% of digital libraries using TF-IDF for this purpose [16].

Owing to the substantial number of misspellings, our corpus contains 865,526 unique words, resulting in high-dimensional TF-IDF vectors for each document. In practice, to alleviate the vector's dimensionality and enhance computational efficiency, data analysts frequently opt to omit word tokens that exhibit a high frequency of occurrence across the corpus, as well as those tokens appearing in only a few documents. The Python library *scikit-learn* offers parameters to facilitate the seamless removal of such tokens [1], a strategy we had to adopt due to computational constraints. It's crucial to highlight that employing an alternate strategy, such as retaining all word tokens at this stage, could potentially impact the analysis results.

We filtered out word tokens that appeared in nine documents or less and word tokens that appeared in more than 80% of the cases, which is a common practice followed by data analysts [119]. The vocabulary size has been reduced to 76,188 and each case is represented by a numerical vector with the same size. The occurrence of words is the most prominent feature encoded in the resulting case representation.

3.2.2 Reducing Feature Dimensions: ***Latent Semantic Analysis (LSA)****.* In the field of Machine Learning (ML), reducing feature dimensions of data is crucial for processing high-dimensional data such as text. Dimensionality refers to the number of features representing a data point, like unique words in documents. Dimensionality reduction techniques are widely employed to address the challenges associated with high-dimensional data, enhancing computational efficiency, facilitating data visualization, and aiding human exploration of text data. However, what often goes unnoticed in this dimensionality reduction process is the substantial loss of information.

Latent Semantic Analysis (LSA) is a popular dimensionality reduction method for text data by modeling topics in a text corpus, i.e., topic modeling [42]. The fundamental assumption is that each case summary has latent semantic structures ("topics") that can be algorithmically identified by listing how frequently words occur in the cases. After identifying these topics, each case can then be represented by assessing the dominance of these latent topics in a case. Baumer et al. [14] explain that in topic modeling the underlying assumption is that "geometric proximity matches human notions of semantic similarity" (p.1399) For example, one would expect the words "Afghanistan" and "Taliban" to appear together in a case file about an asylum seeker from Afghanistan and that an algorithm would identify the same relationship by counting how often these two words occur in the same document. Topics are therefore identified by finding words that often occur together in the case summaries. However, there is also information loss because documents are represented based on their composition of latent topics, while words that don't belong to predominant topics are discarded.

To gain a better understanding of the influence of choices made in the data representation on the outcomes of outlier detection, we experimented with reducing the 76,188-dimensional vectors representing each asylum case into lower-dimensional vectors, using LSA. Determining the ideal number of topics within a corpus, a balance that retains valuable information while eliminating less informative topics, is challenging. Both researchers and practitioners often rely on coherence scores to evaluate the quality of the identified topics. Researchers generally don't agree on the optimal coherence score [56, 94]. We selected three commonly applied coherence scores to obtain a number of dimensions to compare to each other:

— UMASS. This is the most widely used coherence score [68]
— elbow method with accumulated explained variance [124]
— NPMI score [22]

We utilized three distinct scores to assess topic coherence across various dimensions and observed that each score yields a distinct optimal number of topics. Given the common practice, as identified in our literature review, of data analysts visualizing data points in two dimensions to detect outliers, we augmented our set of dimensions accordingly. Consequently, we applied the three ODAs to each resulting dataset, reducing them to dimensions $d = 2, 4, 8, 30$, and enabling a comparative analysis of the labeled outliers across different dimensionalities.

3.3 Phase 2: Applying Outlier Detection Algorithms (ODAs)

In our analysis, we focused on ODAs that enjoy popularity within the field [4, 21]. From this pool, we carefully selected three algorithms that met the specific requirements of our dataset. What sets these algorithms apart are their diverse approaches to calculating a norm. Assigning

Table 2. The Three Applied ODAs and their Characteristics

	LOF	GMM	ABOD
Approach	density-based	clustering-based	angle-based
Scope of Comparison	local	global	global

a data point as an outlier hinges on the algorithm's understanding of what constitutes a *normal* data point. The process of distinguishing whether a data point is normal or an outlier is always calculated relationally, considering its relation to other data points. Various ODAs compute this relation differently, which we will show. Notably, our dataset lacks labeled data, which serves as the foundation for supervised machine learning algorithms. These algorithms rely on labeled data to learn patterns and relationships. In our scenario, the cases could have been labeled as either "asylum denied" or "asylum approved," had it been a labeled dataset. Consequently, we needed to turn to unsupervised outlier detection algorithms. Our choice of these particular algorithms was primarily driven by the unique characteristics of our data, which predominantly comprises textual information and exhibits a high prevalence of misspelled words. Furthermore, the selection of these three distinct variants of ODAs serves the purpose of providing a comprehensive illustration of various outlier detection approaches at work. Table 2 shows an overview of the different categorizations of the used ODAs. During the exploratory phase of our analysis, we experimented with other outlier detection algorithms on our dataset, such as KMeans [64] and COF [112]. For detailed information on the discarded algorithms, please refer to our GitHub repository.

Applying the three ODAs to the four different datasets representing the cases in different dimensions ($d = 2, 4, 8, 30$) results in 12 distinct sets of outliers. In total, 127 data points are designated as outliers by one or more ODA. We are interested in how many of the data points in these 12 sets overlap, which is shown in Figures 2 and 3.

We proceeded to randomly select five instances from the pool of 127 outlier cases using the Python pandas library. This approach enabled us to gain insight into the individuals represented by these data points, thereby shedding light on those who might be excluded from the dataset.

All ODAs we employ fall under the unsupervised machine learning paradigm. In the following sections, we describe how the chosen ODAs produce outliers in two steps:

(1) Calculate a factor indicating its *outlierness*
(2) Identify extreme values in this factor with respect to a given threshold

3.3.1 ***Local Outlier Factor (LOF).*** In 2000, Breunig et al. [25] introduced the LOF algorithm, which has since been widely used to identify outliers in various kinds of data. Many recent ODAs are inspired by the rationale behind LOF and pose mere tweaks to that algorithm [21]. When it was introduced in 2000, this algorithm was novel in the sense that *outlierness* is not considered a binary factor. Instead, it calculates an *outlier score*, based on its local density compared to its neighboring points. In practice that means that the data analyst has two approaches to setting a threshold for norm versus outlier: (1) A data point is considered an outlier, starting from a certain LOF score, meaning that, starting from a certain magnitude of difference in the local densities of the data points, a data point counts as an outlier. This assumes a certain level of knowledge from the analyst to know what constitutes a low LOF score. (2) A certain percentage of the number of data points is defined that are considered outliers. This strategy assumes that the dataset has outliers and that the percentage is known to the data analyst. This assumes either a high domain knowledge of the data analyst or the data analyst uses visualization of the data points in conjunction with experimenting with different percentage values to converge to a value that literally 'seems' reasonable.

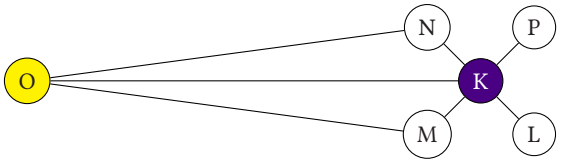


Fig. 1. The intuition behind ABOF. The angles that the outlier O makes with any other point are small and have a low variance. For a point that is close to its neighbors, such as K, the angles vary a lot and will be bigger.

This algorithm measures outlierness by calculating how isolated or distant a point is from its surrounding points. LOF is considered a local outlier detection algorithm (see Table 2), because the outlier score of each data point is calculated using only its k neighboring points and not the whole dataset. k is a parameter that can be set by the data analyst. A low LOF score indicates that the amount of points close to a point is similar to its neighbors and considered normal, while a high LOF score suggests that a point is significantly different and potentially an outlier. This algorithm results in a ranking of all data points.

3.3.2 ***Gaussian Mixture Model (GMM).*** The GMM is a probabilistic clustering model that assumes each data point belongs to a cluster or subpopulation that is normally distributed, i.e., a Gaussian distribution. It is probabilistic because for each data point it is calculated what the probability is that this point belongs to one of the subpopulations.

In the learning phase of this Ml algorithm, the values of three parameters are learned:

— number of subpopulations in the dataset
— mean of each subpopulation
— size of each subpopulation (variance/covariance)

To avoid reaching a *local maximum* in the learning of the clusters, this algorithm is run many times ($n = 100$), and the parameters that best fit the data are chosen. The underlying assumption of this method is that the dataset in its current representation has been generated from a mix of Gaussian distributions with unknown parameters (see list above). Data points that have a low probability to belong to any of the subpopulations learned by the algorithm are considered outliers. The advantage of this method over, e.g., the popular k-means clustering approach, is that it can be used for data where the clusters are not clearly defined and the shapes of the clusters are more flexible. They can be, e.g., elliptical, as opposed to circular. Another advantage is that it is a well-suited algorithm for clusters with different densities. An initial exploration of our data suggests clusters with different densities.

To find an optimal number of clusters, a Bayesian Gaussian mixture model was applied, to calculate an optimal number of subpopulations ($n = 7$).

3.3.3 ***Angle-based Outlier Factor (ABOF).*** ABOF stands for Angle-Based Outlier Factor, an outlier detection algorithm advocated by Kriegel [61]. The fundamental concept behind this method is its reliance on the angles formed between pairs of data points. The underlying idea is that if a data point is significantly distant from the rest of the data, the angles it forms with any other two data points will generally be small and exhibit minimal variation (low variance). This concept is visually demonstrated in Figure 1, specifically at point O. Conversely, if a data point is in close proximity to other data points and is surrounded by them, it is likely to produce a wide range of varying angles with different data points (high variance). This characteristic is illustrated in Figure 1 at point K. To quantify this notion, the ABOF algorithm calculates the angles between each data point and all other points within the dataset, subsequently measuring the variance of these angles. As each data point is compared to every other point, this algorithm is often referred

to as global ODA (see Table 2). Each data point is then assigned an ABOF score based on these angles. Data points characterized by a more limited spectrum of angles will yield lower ABOF values, designating them as potential outliers.

4 Analysis and Findings: Insights from Danish Asylum Case Summaries and Algorithmic Construction

To gain deeper insights into the factors shaping representation within a dataset, our focus is on examining outliers in a dataset of Danish asylum case summaries, as well as the algorithms that construct it. We aim to make data science techniques more tangible by mapping them back onto the reality they aim to model, as a starting point for how we can engage non-data scientists in the decision that are made. We outline the *inner workings* of three ODAs and show how they shape the representation of individual asylum seekers differently by shaping the dataset differently.

In an effort of collective sensemaking [79] with asylum practitioners, we implemented an outlier detection scenario – following best practices in the field – with this dataset and presented our preliminary findings to a caseworker at the Immigration Services. Two of the outlier cases our algorithm identified pertained to individuals who initially were granted asylum but then committed a crime in Denmark, which revoked their right to asylum and a Danish residency. However, they cannot be forced to be sent back to their home country because their asylum motive is still valid, leaving them in a legal grey area and potentially confined to a Danish deportation center for an extended period of time. The response that we received from the caseworkers to these cases was surprise that they would be labeled as outliers and the explanation that there are hundreds of cases like that.

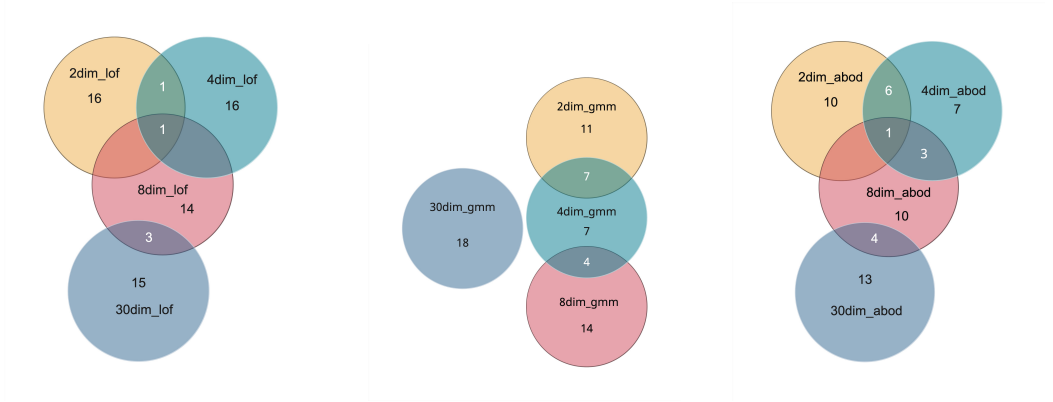
The caseworker's perplexity regarding algorithmically generated outliers prompted us to inquire into the conditions and rationales governing the algorithmic production of outliers and how to make these conditions accessible for discussion and engagement with asylum practitioners. In the following, we contribute findings from our own process of engaging and probing outliers in a dataset of asylum cases.

In the following sections, we present our two key findings: firstly, the substantial impact of data analysts' *featuring* practice and algorithm choices on labeling data points as outliers; and secondly, how algorithms embody and enact concepts such as textual topics and norms, influencing the labeling of outliers in our dataset of asylum decision summaries.

4.1 Finding 1: Data Analyst Choices that have a Significant Influence on the Labeling of Data Points as Outliers

Our first observation and finding reveal a high sensitivity of the categorization of outliers to both the data analysts' featuring practice, as well as the choice of ODA, which we lay out in detail below. Adhering to textbook guidelines [4, 114, 119] for selecting optimal parameters for both cases, we anticipated significant overlap in outlier sets. However, our findings contradict this expectation, highlighting the substantial divergence in the resulting sets of outliers depending on the parameter chosen. The primary influence is the extent of information retained, where higher dimensions correspond to greater retention of information from the case files. For the individual asylum seeker, this means that it is uncertain whether they are represented in a dataset that potentially serves as a ground truth for algorithmic decision-making, and heavily depends on the algorithmic and featuring choices made by data analysts. However, these decisions prioritize metrics like model accuracy rather than considering the representation of individuals.

The inconsistency in outcomes emphasizes that *best practices* for data preparation lack uniformity and frequently hinge on the data analyst's judgment. Furthermore, it implies that the procedural decisions made by data analysts to improve the subsequent performance, such



(a) LOF (b) GMM (c) ABOD

Fig. 2. The three Euler diagrams depicted here illustrate the intersection of outlier sets generated by three distinct ODAs across four dimensions. The black numbers represent the count of cases uniquely belonging to one set of outliers, while the white numbers indicate instances of overlap between sets.

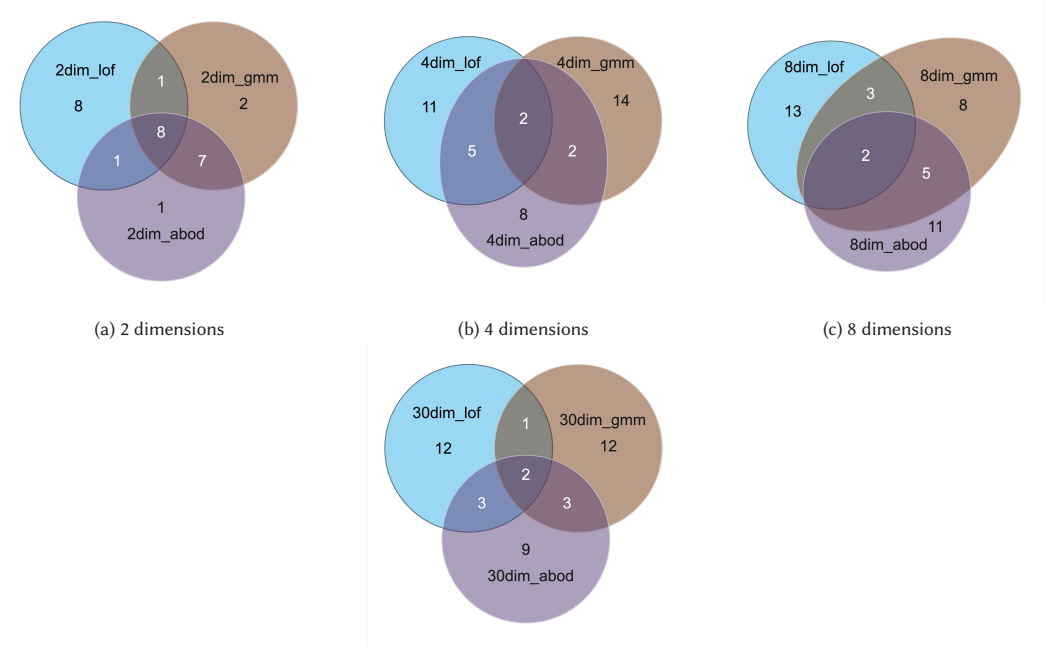
as accuracy of decision algorithms [98, 99], have a substantial effect on which data points are labeled as outliers. These outliers are frequently excluded from the dataset, as discussed in Section 2.3, and consequently impact the outcomes of the decision-making tool. Importantly, these procedural decisions are often made by individuals who may be unaware of their effects and are removed from the actual asylum decision-makers. This indicates that developers of AI tools wield considerable influence over the decision outcomes of the tools they create, yet they are structurally shielded from accountability, as the AI development process is detached from the actual asylum decision-making process (e.g., [74]).

Figures 2 and 3 display the overlap among outlier sets based on two distinct parameters: the selection of ODA and the representation of case files in varying dimensions $d = 2, 4, 8, 30$ using an Euler diagram.[2] A comparison between the two figures reveals that when employing the same ODA but altering the number of data dimensions (Figure 2), there is relatively minimal overlap among outlier sets, particularly evident when utilizing LOF for outlier categorization (Figure 2(a)), in contrast to the intersection of outlier sets across dimensions (Figure 3).

Figure 2 illustrates that as the disparity in data dimensions increases, there is a decrease in overlap among points labeled as outliers. Notably, the least overlap occurs between the outlier sets of the 2- and 4-dimensional datasets, and the 30-dimensional outliers, where there are no overlapping data points for any of the three ODAs. Conversely, a larger overlap is observed between the outlier sets of the 2-dimensional and 4-dimensional datasets across all three ODAs. Figure 3 corroborates this finding, demonstrating that even when employing different ODAs, the greatest overlap exists among outliers categorized in the 2-dimensional dataset (as depicted in Figure 3(a), where 8 data points are labeled as outliers by all three ODAs). Conversely, there is minimal overlap, with only two data points, among the outliers of the 30-dimensional dataset (as shown in Figure 3(d)). Thus, as the amount of information encoded in the data increases (higher dimensionality), the disagreement among ODAs regarding outlier labeling also increases.

To gain insights into the people labeled as outliers in our dataset, we randomly selected five outliers that all represent individual asylum seekers. Table 3 provides an overview and

[2]A Euler diagram is a visual representation that illustrates the intersections between different sets of data using closed curves or shapes, similar to a Venn diagram.



(a) 2 dimensions

(b) 4 dimensions

(c) 8 dimensions

(d) 30 dimensions

Fig. 3. These four Euler diagrams showcase the overlap of outlier sets at varying dimensions in the text representation $d = 2, 4, 8, 30$, comparing the intersection among three ODAs (LOF, GMM, ABOD). The black numbers represent the count of cases uniquely belonging to one set of outliers, while the white numbers indicate instances of overlap between sets.

Table 3. This Table Displays Five Randomly Selected Cases from the Set of 127 Outliers, Indicating the Count of Outlier Detection Algorithms (ODAs) Labeling Each Case as an Outlier for Various Data Dimensions

Case ID	Dimensions: 2	4	8	30	country of origin	decision year	decision	gender	brief description
17186	3	3	0	0	Afghanistan	2020	approved	male	has been appealed at RAB 3 times, original asylum motive changed to converting to Christianity while living in Denmark,
1244	0	1	3	0	Afghanistan	1998	denied	female	several asylum motives: shia-Muslim afraid to be killed by sunni-Muslims, divorced woman, family involved with Taliban opposition party
16615	2	0	0	0	Afghanistan	2019	denied	male	cousin wants to recruit him to Taliban
3138	3	0	0	0	Nigeria	2003	denied	male	convicted for dealing drugs in DK, faces double punishment in Nigeria and DK
10453	0	0	0	2	Burma	2014	denied	male	cannot establish credibility because lived his whole life in Bangladesh, language test can't confirm he's Rohingya, UNHCR records locate family in Burma

Additionally, key characteristics of the cases are provided.

some characteristics of these individuals. The examples affirm that the choice of dimensions significantly influences outlier labeling, exemplified by applicant ID 17186, who was labeled as an outlier in 6 out of 12 datasets, particularly by all ODAs where two and four dimensions are used during the featuring. That means that the more information is retained in the case file about an individual, the more unlikely it is that this person will be categorized as an outlier. Conversely, applicant ID 10453, representing an asylum seeker from Burma, is categorized as an outlier only by two ODAs utilizing 30 dimensions. In this specific instance, this implies that if a data analyst opts to represent each case with 30 dimensions, there is a higher probability of categorizing this asylum seeker as an outlier and consequently eliminating them from the dataset. This action may lead the algorithm to potentially dismiss similar cases due to incomplete information.

4.2 Finding 2: Algorithms Embody and Enact a Distinct Interpretation of Concepts Like Textual Topics and Norms

Delving into the steps of algorithmic calculations leading up to the labeling of outliers, our second finding is that algorithms embody and perform a distinct interpretation of *important latent concepts* in text data, as well as an understanding of *norms* and *normal data points* that we need to pay attention to.

Building on Suchman's caution against treating AI as a monolithic entity and emphasizing the significance of distinct processing steps, the ensuing sections dissect various algorithmic steps employed in outlier detection: turning text into numbers and ODAs. This exploration involves analyzing the retained information of a case when converting case files into numerical representations based on *topics* (Section 4.2.1) and scrutinizing the foundational principles of calculating a *norm* applied to these numerical representations (Section 4.2.2).

4.2.1 Turning Text into Numbers. A critical point of how representation in datsets is shaped is the practice of turning text into numbers. As detailed in Section 3.2.2, the ML technique LSA, also called topic modeling, is often used for dimensionality reduction for text data. It holds the promise of "finding latent and important concepts" [42] within collections of text documents such as asylum case summaries. The *topics* that the algorithm identifies across the corpus of cases are then used to represent cases by assigning to each case how much of a certain topic it contains.

In this section, we explain how the algorithm's notion of *importance* is linked to the variation in the data and how this connection shapes the creation of topics that represent asylum cases. We illustrate that when algorithms are tasked with determining significance and concealed elements of text data, this shapes the computation of outliers and therefore representation in datasets.

LSA identifies latent topics in text data, which are considered *most significant* [42], and represents them as a list of words. Since LSA operates on the form of the text and counts words, it does not guarantee conceptual coherence among the words in a topic. It only ensures their frequent co-occurrence in documents. The words listed in a computational topic therefore often appear conceptually unrelated. To demonstrate this notion, we analyze the two first topics that result from transforming our data into two dimensions (see Table 4). This means that when each case is represented in two dimensions, it is characterized by two numbers denoting the prevalence of the first two topics in the case. By analyzing the words representing the two topics that result from the topic modeling, we learn that different aspects of the cases are represented, such as procedural peculiarities or geographical locations related to an asylum seeker's journey.

plaintiff, the plaintiff, Somalia, Al-Shabaab, Syria, Alien Act, Mogadishu, Lebanon, plaintiff's. In the first latent topic, four of the nine words refer to geographical places, indicating a connection to asylum seekers originating from Afghanistan, Iran, and Iraq. Words like "father" and "mother" also appear, suggesting potential family-related themes in the asylum motive. This could, e.g., be

Table 4. Ten Most Significant Words Representing the Two LSA Topics, Translated from Danish into English

Latent Topic 1	questioned, Afghanistan, plaintiff, Iran, Taliban, Iraq, father, year, mother
Latent Topic 2	plaintiff, the plaintiff, Somalia, Al-Shabaab, Syria, Alien Act, Mogadishu, Lebanon, plaintiff's

that a family member has been killed and the applicant thereafter fled the country. The word "questioned" appears often in the documents, when interview data is included. Sentences in the interview transcripts follow the pattern "When questioned, the applicant replied that...". A high prevalence of this topic dimension in a case indicates one or several interviews.

The second latent topic comprises words associated with Somalia (e.g., Somalia, Al-Shabaab, Mogadishu), as well as other geographical regions such as Syria and Lebanon. In the context of our dataset, this suggests that documents in which this topic is prevalent pertain to asylum seekers originating from these specific countries. The information encoded in the second dimension is more restricted to geographical location and indicates that the applicants originate from Somalia.

In this segment, we demonstrate that condensing our case files into two dimensions (or two numbers) results in encoded information primarily centered around geographical details, family-related themes, and the inclusion and exclusion of interview data in the case files. This highlights a substantial loss of information and nuance inherent in this text encoding approach. We also demonstrate how the algorithm fails to identify *important* concepts using the information content, but rather operates on the form of the data, using word counts and their distribution. Rask Nielsen and Møller [95] underline that crucial aspects of case files, from an asylum decision-maker perspective, include factors like credibility. Previous research on asylum case summaries by legal and data analysts highlighted elements such as asylum motive, country of origin, the perpetrator of persecution (e.g., government, family members, etc.), political involvement of the applicant, sexual orientation, and religion as essential information in an asylum case file [55]. This underscores the divergence of how algorithms and humans determine whether information is to be discarded or retained, which in turn affects the calculation of outliers, emphasizing the need to carefully consider the implications of this algorithmic step of turning text into numbers.

4.2.2 ODAs. The data features used to calculate a *norm* in turn affect the calculation of outliers. In the following, we zoom in on the different interpretations of the concept of the *norm* by each ODA we have applied to demonstrate this point. The clustering-based GMM algorithm assumes data points are grouped into clusters that have internal similarity but differ from points in other clusters. If there is a cluster of data points that deviates from the rest, whether it qualifies as an outlier or as another cluster depends on the quantity of these points, highlighting a cluster-based norm.

By contrast, the LOF algorithm abandons a global norm by computing a norm for each data point based on its local neighborhood. In this localized norm perspective, outliers are identified based on deviations from their immediate surroundings. The ABOF algorithm introduces yet another norm perspective. While it maintains a global notion, it evaluates the isolation of a data point using angles, with low angle variance signifying distance from other points. Defining a norm via angular relationships adds another layer to the diverse interpretations of what constitutes normal or anomalous. Across all these ODAs, *distance* in Euclidean space defines norms and outliers. Normal data points exhibit proximity or similar distances to each other, whereas outliers are located further away. This emphasizes the relational nature of outliers, as their identification is contingent on their relation to other data points, be it neighboring points or the majority in the dataset. Thus, the identification of outliers hinges heavily on the specific collection of data points.

Since all algorithms are based on the distance between data points, we take another look at what the distance in this case stands for. The distance between two documents reflects their similarity or dissimilarity in terms of the topics they contain. If they are close together, documents share similar topics. If the outlier score is derived from these distances, it implies that an outlier is a point with few neighboring points that exhibit a similar prevalence of topics.

This section highlights the different concepts of *norm* embodied and calculated by different ODAs and how individuals are produced as outliers based on these different principles. The Gaussian Mixture Model (GMM) relies on a cluster-based norm, the Local Outlier Factor (LOF) adopts a localized norm perspective based on local neighborhoods, and the Angle-Based Outlier Factor (ABOF) introduces an additional norm perspective based on angles.

We demonstrate that the labeling of individuals as outliers in a dataset relies on how ODAs group individuals as data points into different *shapes* of normal. This in turn determines which asylum seeker is categorized as an outlier in relation to this specific interpretation of *normal* data points.

These findings highlight how the labeling of outliers in asylum case datasets is highly sensitive to the choices made by data analysts, including featuring and the type of outlier detection algorithm used. This crucially depends on the conceptual and statistical definitions of both norms and differences. Algorithms rely on predefined notions of what constitutes *normal* data, with norms being shaped by statistical properties such as clustering patterns or local neighborhoods in multidimensional space. Differences, on the other hand, are calculated based on the distance between data points–whether in terms of textual topics or other features–which dictates how far an individual case diverges from the norm. These definitions of norms and differences are not neutral. They embody specific assumptions about the data and its structure, leading to different interpretations of an outlier. As a result, the representation of asylum seekers in the dataset is not only a reflection of the data itself but also of the algorithmic and conceptual frameworks used to analyze it.

5 Discussion

The need to comprehend the formation of datasets and their influence on training AI models is a question of how we can get closer to what is termed human-centered data science (e.g., [9]). Accuracy is a key concern in this regard and ODAs play a critical role for data analysts to improve the accuracy of their AI models [32, 114]. However, ODAs also play a vital role in determining the inclusion or exclusion of individuals in a dataset that potentially serves as the basis for automated decision tools. This is compounded by the global trend of governments increasingly adopting these tools [43], despite the fact that the decisions mirror historical data, effectively perpetuating past patterns in decision-making [7]. A key concern is that perpetuating past decisions through automated decision tools may reinforce biases and inequities present in historical data, potentially exacerbating existing disparities and hindering progress toward fair and just decision-making. In a work context, such concerns are mitigated through caseworkers' discretion in the decision-making process. In data science, it is crucial to address the risk that these tools may actively shape the future by influencing decision-making processes [7]. By now, it is widely known that algorithmic tools and AI inherently present risks to minorities [18, 23], since machine learning tools are technologies of majorities. They yield a high accuracy for well-represented communities in a training dataset. Yet, in a pluralistic world like ours [45, 67, 97], every dataset has outliers: data points that are unlike the others in one way or another. Taking the outlier as our perspective, as suggested by scholars like Welles [120], we wanted to see what conditions categorize some data points as outliers and others as the norm.

Norms and normativity have traditionally been utilized to marginalize individuals who deviate from the majority, a role that statistics have long played. Given the increasing use of

automated decision tools, it becomes imperative to gain a deeper understanding of algorithmic interpretations of what is considered *normal* and how these mathematical computations shape concepts of normalcy within the various domains, such as asylum decision-making, where these models find application. Increasingly, critical scholars caution about the disconnect between the model's representation of data and the actual domain-specific data [18, 39]. This dichotomy poses the risk of losing sight of what a data model truly represents and the impact it has on the lives of individuals. In this paper, we have demonstrated how representation in datasets is interdependent with (1) data analysts' featuring practice and choice of ODAs, which are removed from the actual decision-making process, as well as (2) the resulting algorithms that enact concepts such as *importance* in textual topics and *norms*.

It is crucial to consider this, because as algorithmic decision-making tools become more prevalent in decision-making processes, they must always be understood within the specific context of their application. The context defines the meaning of each outcome. Asylum decision-making as a domain possesses its unique set of practices with inherent complexities, such as, e.g., the assessment of the credibility of the applicant [95]. The algorithmic tools and AI used in this context reduce the openings for human judgment and compassion that can happen in a situated manner [71], yet human judgment is introduced in other stages of the decision process, such as the data cleaning phase of the AI tool development [73, 74, 87, 99, 123]. Louise Amoore has elucidated how algorithms' probabilistic outputs are transformed and put into action, serving as the basis for decisions that ultimately influence and mold the future [7]. In our case, we took a step back to understand what comes before an algorithm calculates a prediction for a decision outcome, for example, a probability for a particular individual. We examined how who is represented (i.e., included) in a dataset is shaped by the algorithmic tools that are applied. This is important for at least two reasons. First, the quality (coverage) of the algorithm suffers when the data becomes impoverished with respect to diversity. Second, exclusion from the dataset imposes direct harm on the excluded humans and their families. Just as the nuanced choices made by AI developers shape outlier categorization, they also dictate the ramifications of being labeled as such. AI developers can adopt different strategies in the handling of outliers: opting for default decisions (e.g., denying an asylum claim), communicating decision uncertainty for individuals or communities with insufficient algorithmic knowledge, or imputing missing individual information based on normative data from the training dataset.

A question for this research is how we can build a body of work that opens up this discussion to a broader audience to address the colonialism of data science. Data colonialism refers to a concept that parallels historical colonial practices with modern data practices [35]. It underscores issues such as the perpetuation of existing power imbalances through data-driven technologies and practices, often exploiting and controlling data from marginalized communities with regard to aspects like race and gender [67]. For many, using data science tools for public decision-making promises more efficiency and to mitigate human biases. But the often opaque algorithmic tools don't erase, but merely shift the power imbalances and possible human biases, leading to the perpetuation of the colonialism of data science, in which data analysts wield power over those whose data is processed, and possibly marginalize certain groups and communities, that are not captured by conventional categories such as race or gender [52].

In our analysis of asylum cases, we found, similar to what Amoore argues, that algorithmic tools and AI shape outcomes by determining representation of individuals and communities in datasets. Notably, early stages of AI tool development, including data cleaning, emerge as integral components of casework. This holds true not only when caseworkers utilize automated decision tools, but also when they themselves generate case files, engaging in data work [70, 83] and shaping the information recorded in a dataset. The everyday practice of taking decisions therefore shapes

the data cleaning process of a data science project, including outlier handling where it is decided who is included and excluded in a dataset. Consequently, outliers play a crucial role in shaping decision outcomes, particularly in data-driven decision-making, where datasets, like the one in focus for this study, are often treated as ground truth [76, 123]. With asylum decision-making increasingly relying on data, human beings are being represented as data points, necessitating a careful examination of the mechanisms and dynamics involved in algorithmic processing.

Nonetheless, the responsibility of evaluating the consequences of algorithmic tools cannot rest solely on data scientists, as the network of power and decision-making is far more complex [33]. Decisions about implementing these tools and defining their requirements often involve a broader set of stakeholders, including teams, supervisors, lawmakers, and institutions [75, 76, 101, 123]. As highlighted by Sambasivan et al. [99] and Muller and Strohmayer [74], data analysts frequently overlook the downstream ramifications of seemingly minor decisions, such as parameter choices or algorithm selection for data cleaning. These decisions are typically made with a focus on optimizing the accuracy of the data model, meaning how accurately a data model describes a ground truth dataset, not taking into account, that these datasets are often already *cleaned* of outliers, thus failing to represent the diverse range of cases and people encountered by asylum practitioners. Consequently, data analysts optimize with regard to the dataset rather than the individuals affected by their developed tools.

We propose two entry points for discussing algorithmic tools and their application in the data cleaning process. We demonstrate that the construction of outliers is determined by an interplay between (1) algorithmic enactment of abstract concepts, and (2) data analysts' choices:

(1) **Algorithmic steps enact abstract concepts.** Algorithms incorporate and manifest abstract concepts like importance and norm, translating them into mathematical calculations that may lack human intuition. As demonstrated in this paper, algorithms engage in quantitative manipulation of formal patterns rather than semantic interrogation, emphasizing the need to scrutinize how concepts like norm and normativity, or topics that represent important topics in case files, are operationalized within formal manipulations.

(2) **Data analysts' choices during data cleaning influence the resulting outcome.** We demonstrate that the discretionary decisions of data analysts regarding parameter selection, even within the framework of best practices, significantly impact the categorization of individuals as outliers. Our findings indicate that seemingly minor choices during the data cleaning process, like selecting a metric to determine data dimensionality or choosing an ODA, wield considerable influence over who is labeled as an outlier. As suggested by Rieder [98] and Sambasivan et al. [99], these procedural decisions are frequently geared towards optimizing downstream factors such as the accuracy of applied data or predictive models, often overlooking the consideration of the human individuals represented in the data.

We show how to critically engage with outlier handling as a significant juncture that molds representation within datasets. We illustrate how to scrutinize procedural decisions made by data analysts during data cleaning using the example of outlier production, as well as the algorithmic steps preceding the application of a predictive model when evaluating an algorithm. Thus, we provide insights toward moving beyond the prevalent practice of treating algorithms as a monolithic entity [109] and instead analyzing algorithms as a sociotechnical entity [103] consisting of several processing steps.

The insights presented in this paper should give rise to concern, but also incite interest around an opportunity to expand our toolbox for auditing algorithms. Traditional algorithmic auditing techniques commonly depend on assessing the bias of a specific algorithm within a distinct

domain towards particular demographic groups, as highlighted by Bandy [11]. This approach necessitates an initial understanding of which demographic groups may be pertinent, often aligning with protected categories in many countries, such as gender, ethnicity, race, and physical ability, as indicated by Bender et al. [18]. Affirming this perspective, boyd [23] emphasizes that algorithmic systems carry an even greater risk of introducing novel forms of discrimination not captured by these conventional categories. boyd contends that critical researchers must be more aware of other forms of discrimination that are not captured in categories that are "already dominant in socio-political consciousness" (p. 7). Using different ODAs and data representation parameters can help identify these new forms and categories of discrimination that are possibly introduced by algorithmic decision tools. Our study illustrates how algorithms prioritize distinct aspects of case files, such as geographic details and the presence of interview data, in contrast to human judgment who find different aspects of a case file *important*. This variance in information selection can assist in identifying potentially marginalized individuals and groups beyond conventional protected categories, such as race or gender.

For large-institution work, there is often a trade-off or tension between efficiency and quality. When faced with many cases, efficiency may be important, to take care of as many people as possible. When faced with diverse cases, quality may be important to work with the unique situations of the people in each case. People who favor efficiency often prefer large-scale algorithmic solutions, with decision-making power vested in those algorithms. People who favor quality often prefer per-case considerations, with decision-making power vested in case workers. Organizations often struggle with these potentially contradictory priorities. This is more importantly (perhaps) due to the nature of democratic society, where legal rules and regulations constantly evolve and therefore need interpretation when put into practice. We propose a reconsideration of these decisions in terms of three major factors: technology, policies, and work practices. Taken to an extreme, the efficiency-oriented approaches focus on technology (algorithms) as decisive, with policies that limit the discretion of caseworkers and their expertise expressed through work practices. Taken to a contrasting extreme, the quality-oriented approaches focus on case workers as decisive, with policies that limit consistency or certain forms of "fairness" across diverse cases. However, we note that there are multiple definitions of fairness [69, 102, 117], with potentially poor mapping of algorithmic concepts onto legal concepts [121]. Furthermore, Green [53] has argued that it is formally impossible to satisfy all fairness criteria simultaneously. Thereby, we suggest that human actors need some flexibility in how they apply these three sources of decision-making - technology, policies, and work practices - to specific cases. Any single source of decision authority is likely to be unfair in at least some cases - and differently unfair from one case to another. Fairness in outlier assignment, and in refugee outcomes, remains a difficult problem without a single formal method for making decisions. Policies can be adjusted to allow various latitudes of human decision-making practices (e.g., on the basis of experience, skills, certifications, or seniority). Humans have encoded forms of wisdom in all three sources of decision-making: When the outcomes can benefit or harm humans, then our interim proposal is to trust human case workers to make good decisions based on all three sources.

Moving forward, our study aims to highlight the data points labeled as outliers within our dataset and initiate discussions with Danish asylum practitioners regarding our findings. We aim to understand the unique characteristics that render these cases as outliers in the eye of our applied algorithms and explore how the Danish asylum process addresses such instances. We are also keen on understanding how asylum practitioners define an outlier and how their perspective differs from the algorithmic viewpoint. Given the sensitive nature of asylum decision-making, increasingly influenced by data-driven technologies, it is imperative to scrutinize potential discriminatory impacts on a population already in a vulnerable position.

Acknowledgments

We want to express our sincere gratitude to Prof. Irina Shklovski for her thoughtful feedback on an early draft of this paper. We also thank Colleen Jankovic for her careful editing during the final stages of the manuscript. Additionally, we are deeply grateful to the anonymous reviewers for their constructive comments and insightful suggestions, which have greatly enhanced the quality of this work. Finally, we acknowledge and thank Nordforsk for their financial support through the Nordasil project, which made this research possible.

References

[1] sklearn.TfidfVectorizer. 2024. https://scikit-learn/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

[2] 2019. *National Strategi for Kunstig Intelligens*. Finansministeriet and Erhvervsministeriet, Dansk Regering. https://digst.dk/media/19302/national_strategi_for_kunstig_intelligens_final.pdf

[3] Veronica Abebe, Gagik Amaryan, Marina Beshai, Ilene, Ali Ekin Gurgen, Wendy Ho, Naaji R. Hylton, Daniel Kim, Christy Lee, Carina Lewandowski, et al. 2022. Anti-racist HCI: Notes on an emerging critical technical practice. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–12.

[4] Charu C. Aggarwal. 2017. *Outlier Analysis*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-47578-3

[5] Marwan Alakhras, Mousa Hussein, and Mourad Oussalah. 2020. Location fixing and fingerprint matching fingerprint map construction for indoor localization. *Journal of Sensors* 2020 (2020), 1–14.

[6] Ali Alkhatib and Michael Bernstein. 2019. Street-level algorithms: A theory at the gaps between policy and decisions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[7] Louise Amoore. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press. http://www.jstor.org/stable/j.ctv11g97wm

[8] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica* (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[9] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. 2022. *Human-Centered Data Science - An Introduction*. MIT Press.

[10] Otmane Azeroual. 2020. Data wrangling in database systems: Purging of dirty data. *Data* 5, 2 (2020). https://doi.org/10.3390/data5020050

[11] Jack Bandy. 2021. Problematic machine behavior: A systematic literature review of algorithm audits. (Feb. 2021). http://arxiv.org/abs/2102.04256. arXiv:2102.04256 [cs].

[12] Luye Bao, Nicole M. Krause, Mikhaila N. Calice, Dietram A. Scheufele, Christopher D. Wirz, Dominique Brossard, Todd P. Newman, and Michael A. Xenos. 2022. Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior* 130 (2022), 107182.

[13] Shaowen Bardzell. 2010. Feminist HCI: Taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1301–1310.

[14] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (June 2017), 1397–1410. https://doi.org/10.1002/asi.23786

[15] Anja Bechmann. 2019. Data as humans: Representation, accountability, and equality in big data. *Human Rights in the Age of Platforms* (2019). https://doi.org/10.7551/mitpress/11304.003.0008

[16] Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. 2016. Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries* 17, 4 (2016), 305–338. https://doi.org/10.1007/s00799-015-0156-0

[17] Hanne Beirens. 2022. Rebooting the asylum system. *The Role of Digital Tools in International Protection* (2022). Washington, DC: Migration Policy Institute.

[18] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT'21)*. Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

[19] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code* (1st edition). Polity, Medford, MA.

[20] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. 2021. Envisioning communities: A participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 425–436.

[21] Azzedine Boukerche, Lining Zheng, and Omar Alfandi. 2020. Outlier detection: Methods, models, and classification. *Comput. Surveys* 53, 3 (June 2020), 55:1–55:37. https://doi.org/10.1145/3381028
[22] Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of the Biennial GSCL Conference 2009* (01 2009).
[23] danah boyd. 2023. The structuring work of algorithms. *Daedalus* 152 (02 2023), 236–240. https://doi.org/10.1162/daed_a_01983
[24] danah boyd and Kate Crawford. 2012. Critical questions for big data. *Information, Communication & Society* 15, 5 (2012), 662–679. https://doi.org/10.1080/1369118X.2012.678878
[25] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. LOF: Identifying density-based local outliers. *SIGMOD Rec.* 29, 2 (2000), 93–104. https://doi.org/10.1145/335191.335388
[26] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)*, Sorelle A. Friedler and Christo Wilson (Eds.), Vol. 81. PMLR, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html
[27] Federico Cabitza, Caterina Fregosi, Andrea Campagner, and Chiara Natali. 2024. Explanations considered harmful: The impact of misleading explanations on accuracy in hybrid human-AI decision making. In *Explainable Artificial Intelligence*, Luca Longo, Sebastian Lapuschkin, and Christin Seifert (Eds.). Springer Nature Switzerland, Cham, 255–269.
[28] Hilary Evans Cameron, Avi Goldfarb, and Leah Morris. 2021. Artificial intelligence for a reduction of false denials in refugee claims. *Journal of Refugee Studies* (2021).
[29] Oswald Campesato. 2022. *Data Wrangling Using Pandas, SQL, and Java*. Mercury Learning and Information, Berlin, Boston. https://doi.org/doi:10.1515/9781683929031
[30] Ander Carreño, Iñaki Inza, and Jose A. Lozano. 2020. Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework. *Artificial Intelligence Review* 53 (2020), 3575–3594.
[31] Daniel L. Chen and Jess Eagel. 2017. Can machine learning help predict the outcome of asylum adjudications? *Proceedings of ACM ICAIL Conference* (2017). https://doi.org/10.475/123_4
[32] Xu Chu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. 2016. Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 International Conference on Management of Data (SIGMOD'16)*. Association for Computing Machinery, New York, NY, USA, 2201–2206. https://doi.org/10.1145/2882903.2912574
[33] Patricia Hill Collins. 1990. Black feminist thought in the matrix of domination. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* 138, 1990 (1990), 221–238.
[34] Sasha Costanza-Chock. 2020. *Design Justice: Community-led Practices to Build the Worlds We Need*. The MIT Press.
[35] Nick Couldry and Ulises A. Mejias. 2019. Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20, 4 (2019), 336–349. https://doi.org/10.1177/1527476418796632
[36] Jay Cunningham, Gabrielle Benabdallah, Daniela Rosner, and Alex Taylor. 2023. On the grounds of solutionism: Ontologies of Blackness and HCI. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–17.
[37] Nana Kesewaa Dankwa and Claude Draude. 2021. Setting diversity at the core of HCI. In *International Conference on Human-Computer Interaction*. Springer, 39–52.
[38] Shiloh Deitz. 2023. Outlier bias: AI classification of curb ramps, outliers, and context. *Big Data & Society* 10, 2 (2023), 20539517231203669. https://doi.org/10.1177/20539517231203669
[39] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. MIT Press, Cambridge, MA, USA.
[40] Paul Dourish. 2016. Algorithms and their others: Algorithmic culture in context. *Big Data & Society* 3, 2 (Dec. 2016), 2053951716665128. https://doi.org/10.1177/2053951716665128. Publisher: SAGE Publications Ltd.
[41] Stefania Druga, Fee Lia Christoph, and Amy J. Ko. 2022. Family as a third space for AI literacies: How do children and parents learn about AI together?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.
[42] Susan Dumais, George Furnas, Thomas Landauer, Scott Deerwester, and Richard Harshman. 1996. Using latent semantic analysis to improve access to textual information. *Proceedings, CHI* 88 (01 1996). https://doi.org/10.1145/57167.57214
[43] Matt Dunn, Levent Sagun, Hale Şirin, and Daniel Chen. 2017. Early predictability of asylum court decisions. In *Proceedings of the 16th Edition of the International Conference on Articial Intelligence and Law (ICAIL'17)*. Association for Computing Machinery, New York, NY, USA, 233–236. https://doi.org/10.1145/3086512.3086537
[44] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The algorithmic imprint. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1305–1317.
[45] Arturo Escobar. 2018. *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press.

[46] Melanie Feinberg. 2017. A design perspective on data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 2952–2963. https://doi.org/10.1145/3025453.3025837

[47] Melanie Feinberg, Daniel Carter, and Julia Bullard. 2014. A story without end: Writing the residual into descriptive infrastructure. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS'14)*. Association for Computing Machinery, New York, NY, USA, 385–394. https://doi.org/10.1145/2598510.2598553

[48] Emilio Ferrara. 2023. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci* 6, 1 (2023), 3.

[49] Xavier Ferrer, Tom van Nuenen, Jose M. Such, Mark Coté, and Natalia Criado. 2021. Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine* 40, 2 (2021), 72–80.

[50] Daniel Ghezelbash, Keyvan Dorostkar, and Shannon Walsh. 2022. A data driven approach to evaluating and improving judicial decision-making: Statistical analysis of the judicial review of refugee cases in Australia. *The University of New South Wales Law Journal* 45, 3 (2022), 1085–1123.

[51] Rubén González-Sendino, Emilio Serrano, Javier Bajo, and Paulo Novais. 2023. A review of bias and fairness in artificial intelligence. (2023).

[52] Catriona Gray. 2023. More than extraction: Rethinking data's colonial political economy. *International Political Sociology* 17, 2 (2023), olad007.

[53] Ben Green. 2022. Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology* 35, 4 (2022), 90.

[54] M. E. Haese. 2021. *Modernizing Decision-making: A Comparison of Algorithmic Decision Support Systems in Child Protective Service Agencies in Germany and Australia*. B.S. thesis. University of Twente.

[55] William Hamilton Byrne, Thomas Gammeltoft-Hansen, Sebastiano Piccolo, Naja Holten Møller, Tijs Slaats, and Panagiota Katsikouli. 2023. Data-driven futures of international refugee law. *Journal of Refugee Studies* (02 2023), feac069. https://doi.org/10.1093/jrs/feac069.

[56] Alexander Hoyle, Pranav Goel, Denis Peskov, Andrew Hian-Cheong, Jordan Boyd-Graber, and Philip Resnik. 2021. Is automated topic model evaluation broken?: The incoherence of coherence. (2021). arXiv:cs.CL/2107.02173

[57] Kristin Kaltenhäuser, Tijs Slaats, Thomas Gammeltoft-Hansen, and Naja Holten Møller. 2022. Deconstructing gender in asylum categories: An archival perspective on a practice with limited access. *Proceedings of the 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies* (2022). https://doi.org/10.48340/ecscw2022_n03

[58] Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. 2011. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization* 10, 4 (2011), 271–288. https://doi.org/10.1177/1473871611415994

[59] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving human-AI partnerships in child welfare: Understanding worker practices, challenges, and desires for algorithmic decision support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[60] Anna Kawakami, Venkatesh Sivaraman, Logan Stapleton, Hao-Fei Cheng, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. "Why Do I Care What's Similar?" Probing challenges in AI-assisted child welfare decision-making through worker-AI interface design concepts. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 454–470.

[61] Hans-Peter Kriegel, Matthias Schubert, and Arthur Zimek. 2008. Angle-based outlier detection in high-dimensional data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*. Association for Computing Machinery, New York, NY, USA, 444–452. https://doi.org/10.1145/1401890.1401946

[62] Min Kyung Lee, Nina Grgić-Hlača, Michael Carl Tschantz, Reuben Binns, Adrian Weller, Michelle Carney, and Kori Inkpen. 2020. Human-centered approaches to fair and responsible AI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[63] Xiaohui Liu, Gongxian Cheng, and John Xingwang Wu. 2002. Analyzing outliers cautiously. *IEEE Transactions on Knowledge and Data Engineering* 14, 2 (2002), 432–437.

[64] J. B. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. University of California Press, 281–297.

[65] Helena M. Mentis, Ahmed Rahim, and Pierre Theodore. 2016. Crafting the image in surgical telemedicine. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 744–755.

[66] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between subjectivity and imposition: Power dynamics in data annotation for computer vision. (2020). arXiv:cs.HC/2007.14886

[67] Stefania Milan and Emiliano Treré. 2019. Big Data from the South(s): Beyond data universalism. *Television & New Media* 20, 4 (2019), 319–335.

[68] David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Regina Barzilay and Mark Johnson (Eds.). Association for Computational Linguistics, Edinburgh, Scotland, UK., 262–272. https://aclanthology.org/D11-1024

[69] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application* 8, 1 (2021), 141–163.

[70] Naja Holten Møller, Claus Bossen, Kathleen H. Pine, Trine Rask Nielsen, and Gina Neff. 2020. Who does the work of data? *Interactions* 27, 3 (2020), 52–55.

[71] Naja Holten Møller, Irina Shklovski, and Thomas T. Hildebrandt. 2020. Shifting concepts of value: Designing algorithmic decision-support systems for public services. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI'20)*. Association for Computing Machinery, New York, NY, USA, Article 70, 12 pages. https://doi.org/10.1145/3419249.3420149

[72] Michael Muller, Cecilia Aragon, Shion Guha, Marina Kogan, Gina Neff, Cathrine Seidelin, Katie Shilton, and Anissa Tanweer. 2020. Interrogating data science. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 467–473.

[73] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300356

[74] Michael Muller and Angelika Strohmayer. 2022. Forgetting practices in the data sciences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI'22)*. Association for Computing Machinery, New York, NY, USA, Article 323, 19 pages. https://doi.org/10.1145/3491102.3517644

[75] Michael Muller and Justin Weisz. 2022. Frameworks for collaborating humans and AIs: Sequence and sociality in organizational applications. In *CHIWORK*.

[76] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, et al.. 2021. Designing ground truth and the social life of labels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI'21)*. Association for Computing Machinery, New York, NY, USA, Article 94, 16 pages. https://doi.org/10.1145/3411764.3445402

[77] Christopher J. L. Murray. 2022. The global burden of disease study at 30 years. *Nature Medicine* 28, 10 (2022), 2019–2026.

[78] Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. 2018. Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* 21, 7 (2018), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

[79] Naja Holten Møller, Gina Neff, Jakob Grue Simonsen, Jonas Christoffer Villumsen, and Pernille Bjørn. 2021. Can workplace tracking ever empower? Collective sensemaking for the responsible use of sensor data at work. *Proceedings of the ACM on Human-Computer Interaction* 5, GROUP (July 2021), 219:1–219:21. https://doi.org/10.1145/3463931

[80] Keziah Naggita and J. Ceasar Aguma. 2022. The Equity Framework: Fairness beyond equalized predictive outcomes. (2022). arXiv:cs.CY/2205.01072

[81] Jack Nicholls, Aditya Kuppa, and Nhien-An Le-Khac. 2021. Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. *IEEE Access* 9 (2021), 163965–163986.

[82] Trine Rask Nielsen, Thomas Gammeltoft-Hansen, and Naja Holten Møller. 2024. Mobile phone data transforming casework in asylum decision-making: Insights from the Danish case. *ACM J. Responsib. Comput.* 1, 4, Article 27 (Dec. 2024), 29 pages. https://doi.org/10.1145/3696469

[83] Trine Rask Nielsen, Maria Menendez-Blanco, and Naja Holten Møller. 2023. Who cares about data? Ambivalence, translation, and attentiveness in asylum casework. *Computer Supported Cooperative Work (CSCW)* 32, 4 (2023), 861–910.

[84] Ihudiya Finda Ogbonnaya-Ogburu, Angela D. R. Smith, Alexandra To, and Kentaro Toyama. 2020. Critical race theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[85] Cheong Hee Park. 2023. A comparative study for outlier detection methods in high dimensional text data. *Journal of Artificial Intelligence and Soft Computing Research* 13, 1 (Jan. 2023), 5–17. https://doi.org/10.2478/jaiscr-2023-0001

[86] Samir Passi and Solon Barocas. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*'19)*. Association for Computing Machinery, New York, NY, USA, 39–48. https://doi.org/10.1145/3287560.3287567

[87] Samir Passi and Steven J. Jackson. 2018. Trust in data science: Collaboration, translation, and accountability in corporate data science projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 136 (Nov. 2018), 28 pages. https://doi.org/10.1145/3274405

[88] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (Nov. 2021), 100336. https://doi.org/10.1016/j.patter.2021.100336

[89] Sebastiano Antonio Piccolo, Panagiota Katsikouli, Thomas Gammeltoft-Hansen, and Tijs Slaats. 2023. On predicting and explaining asylum adjudication. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL'23)*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/3594536.3595155

[90] Kathleen Pine, Clauss Bossen, Naja Holten Møller, Milagros Miceli, Alex Jiahong Lu, Yunan Chen, Leah Horgan, Zhaoyuan Su, Gina Neff, and Melissa Mazmanian. 2022. Investigating data work across domains: New perspectives on the work of creating data. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[91] Kathleen H. Pine and Max Liboiron. 2015. The politics of measurement and action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 3147–3156. https://doi.org/10.1145/2702123.2702298

[92] Thomas V. Pollet and Leander van der Meij. 2017. To remove or not to remove: The impact of outlier handling on significance testing in testosterone data. *Adaptive Human Behavior and Physiology* 3, 1 (March 2017), 43–60. https://doi.org/10.1007/s40750-016-0050-z

[93] Adam Ewell Quinn. 2016. *Predictors of Secondary Traumatic Stress among Clinical Social Workers: A Focus on the Impact of the Supervisory Relationship*. Ph.D. Dissertation. University of Georgia.

[94] Hamed Rahimi, Jacob Louis Hoover, David Mimno, Hubert Naacke, Camelia Constantin, and Bernd Amann. 2023. Contextualized topic coherence metrics. (2023). arXiv:cs.CL/2305.14587

[95] Trine Rask Nielsen and Naja Holten Møller. 2022. Data as a lens for understanding what constitutes credibility in asylum decision-making. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP, Article 6 (Jan. 2022), 23 pages. https://doi.org/10.1145/3492825

[96] Tye Rattenbury, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, and Connor Carreras. 2017. *Principles of Data Wrangling: Practical Techniques for Data Preparation*. O'Reilly Media, Inc.

[97] Bernd Reiter. 2018. *Constructing the Pluriverse: The Geopolitics of Knowledge*. Duke University Press.

[98] Bernhard Rieder. 2017. Scrutinizing an algorithmic technique: The Bayes classifier as interested reading of reality. *Information, Communication & Society* 20, 1 (Jan. 2017), 100–117. https://doi.org/10.1080/1369118X.2016.1181195. Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2016.1181195

[99] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M. Aroyo. 2021. Data cascades in high-stakes AI: Everyone wants to do the model work, not the data work. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI'21)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445518

[100] 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.

[101] Devansh Saxena and Shion Guha. 2024. Algorithmic harms in child welfare: Uncertainties in practice, organization, and street-level decision-making. *ACM Journal on Responsible Computing* 1, 1 (2024), 1–32.

[102] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. 2019. How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 99–106.

[103] Nick Seaver. 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4, 2 (Dec. 2017), 205395171773810. https://doi.org/10.1177/2053951717738104

[104] Rifat Ara Shams, Didar Zowghi, and Muneera Bano. 2023. AI and the quest for diversity and inclusion: A systematic literature review. *AI and Ethics* (2023), 1–28.

[105] Ben Shneiderman. 2022. *Human-centered AI*. Oxford University Press.

[106] Nathalie A. Smuha. 2021. Beyond the individual: Governing AI's societal harm. *Internet Policy Review* 10, 3 (2021).

[107] Katta Spiel. 2021. "Why are they all obsessed with gender?"–(Non) binary navigations through technological infrastructures. In *Designing Interactive Systems Conference 2021*. 478–494.

[108] Angelika Strohmayer and Michael Muller. 2023. Data-ing and Un-Data-ing. *Interactions* 30, 3 (May 2023), 38–42. https://doi.org/10.1145/3587240

[109] Lucy Suchman. 2023. The uncontroversial 'thingness' of AI. *Big Data & Society* 10, 2 (July 2023), 20539517231206794. https://doi.org/10.1177/20539517231206794. SAGE Publications Ltd.

[110] Harini Suresh and John V. Guttag. 2021. A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–9. https://doi.org/10.1145/3465416.3483305. arXiv:1901.10002 [cs, stat].

[111] Ki Hyun Tae, Yuji Roh, Young Hun Oh, Hyunsu Kim, and Steven Euijong Whang. 2019. Data cleaning for accurate, fair, and robust models: A big data - AI integration approach. In *Proceedings of the 3rd International Workshop on*

Data Management for End-to-End Machine Learning (DEEM'19). Association for Computing Machinery, New York, NY, USA, Article 5, 4 pages. https://doi.org/10.1145/3329486.3329493

[112] Jian Tang, Zhixiang Chen, Ada Wai-Chee Fu, and David W. Cheung. 2002. Enhancing effectiveness of outlier detections for low density patterns. In *Advances in Knowledge Discovery and Data Mining*, Ming-Syan Chen, Philip S. Yu, and Bing Liu (Eds.). Springer Berlin, Berlin, 535–548.

[113] Anissa Tanweer, Cecilia R. Aragon, Michael Muller, Shion Guha, Samir Passi, Gina Neff, and Marina Kogan. 2022. Interrogating human-centered data science: Taking stock of opportunities and limitations. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[114] Sergios Theodoridis and Konstantinos Koutroumbas. 2009. *Pattern Recognition* (4th ed.). Elsevier/Academic Press, Amsterdam.

[115] Nanna Bonde Thylstrup, Daniela Agostinho, Annie Ring, Catherine D'Ignazio, and Kristin Veel (Eds.). 2021. *Uncertain Archives: Critical Keywords for Big Data*. MIT Press, Cambridge, MA, USA.

[116] Jutta Treviranus. 2014. The value of the statistically insignificant. *Educause* January/February 2014 (01 2014), 46.

[117] Sahil Verma and Julia Rubin. 2018. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness*. 1–7.

[118] Hongzhi Wang, Mohamed Jaward Bah, and Mohamed Hammad. 2019. Progress in outlier detection techniques: A survey. *IEEE Access* 7 (2019), 107964–108000. https://doi.org/10.1109/ACCESS.2019.2932769. IEEE Access.

[119] Sholom M. Weiss, Nitin Indurkhya, Tong Zhang, and Fred J. Damerau. 2005. *Text Mining*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-34555-0

[120] Brooke Foucault Welles. 2014. On minorities and outliers: The case for making Big Data small. *Big Data & Society* 1, 1 (April 2014). https://doi.org/10.1177/2053951714540613. SAGE Publications Ltd.

[121] Alice Xiang and Inioluwa Deborah Raji. 2019. On the legal compatibility of fairness definitions. *arXiv preprint arXiv:1912.00761* (2019).

[122] Jaime Yeckle and Bo Tang. 2018. Detection of electricity theft in customer consumption using outlier detection algorithms. In *2018 1st International Conference on Data Intelligence and Security (ICDIS'18)*. IEEE, 135–140.

[123] Hubert Dariusz Zając, Natalia Rozalia Avlona, Finn Kensing, Tariq Osman Andersen, and Irina Shklovski. 2023. Ground truth or dare: Factors affecting the creation of medical datasets for training AI. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES'23)*. Association for Computing Machinery, New York, NY, USA, 351–362. https://doi.org/10.1145/3600211.3604766

[124] Enes Zvornicanin. 2021. When coherence score is good or bad in topic modeling? | Baeldung on Computer Science. (Dec. 2021). https://www.baeldung.com/cs/topic-modeling-coherence-score

Received 3 May 2024; revised 23 September 2024; accepted 2 December 2024