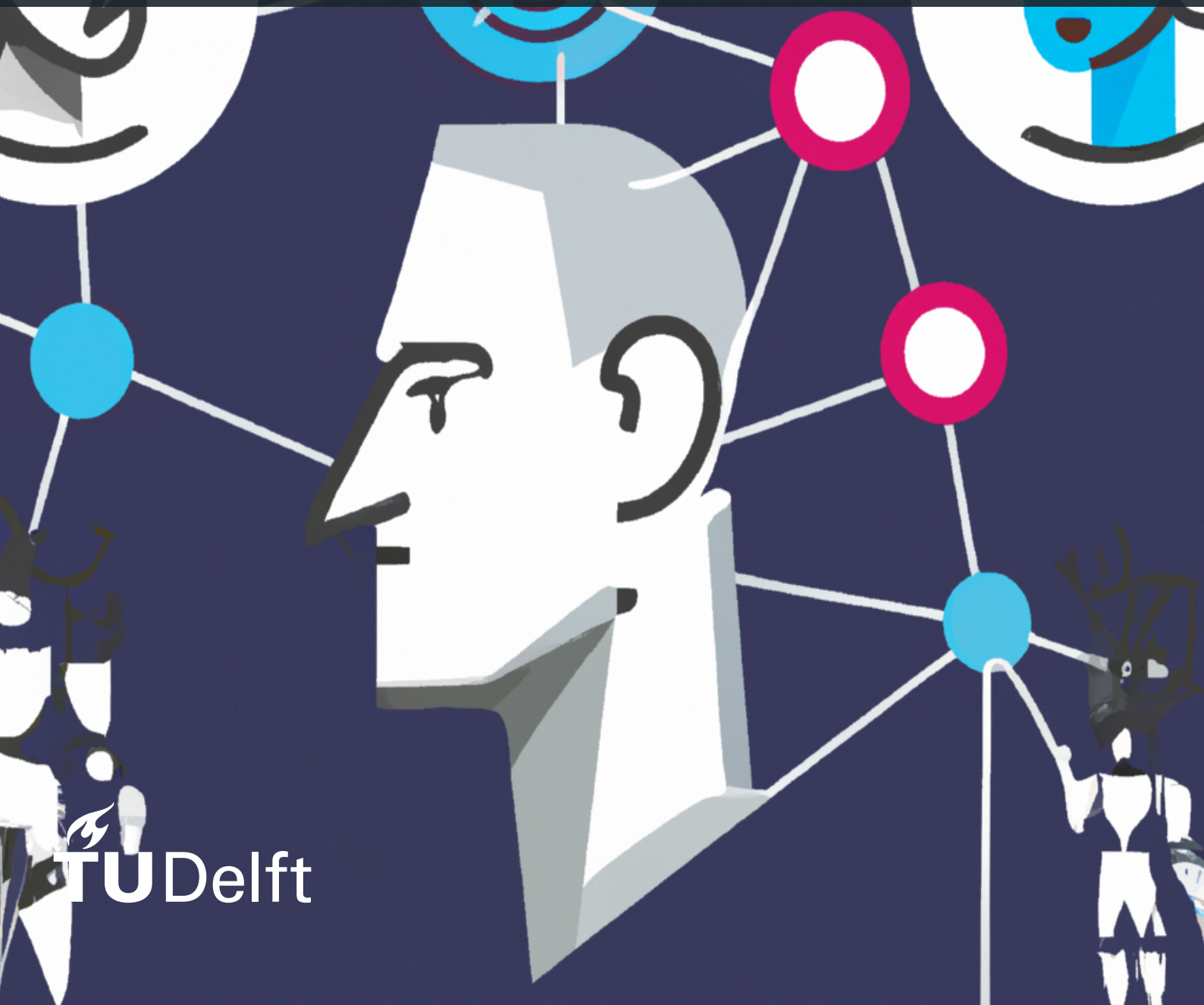
Participatory AI in marginalized communities

Exploring Strategies for Inclusive Stakeholder Engagement in Algorithmic Development

Chakir el Moussaoui



TU Delft

Participatory AI in marginalized communities

Exploring Strategies for Inclusive Stakeholder Engagement in Algorithmic Development

by

Chakir el Moussaoui

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly in June, 2023

Program track:	Data Science & Technology
Thesis committee:	Dr. C.C.S Liem (Supervisor and Chair)
	Dr. L.C Siebert

Preface

It is with immense pleasure and a profound sense of accomplishment that I present this Master thesis to you.

This research endeavor has been made possible through the invaluable guidance, intellectual rigor, and constructive feedback that Dr. Cynthia Liem has provided me with throughout this research. I am truly fortunate to have had the opportunity to work under her supervision and benefit from her wealth of knowledge and expertise. I would like to thank Dr. L. Cavalcante Siebert for accepting to be part of the thesis committee.

I would also like to express my sincere appreciation to my family and friends whose unwavering support, encouragement, and belief in my abilities have been a constant source of motivation throughout this research journey. In particular, the understanding patience, and unconditional support of my parents have been instrumental in overcoming challenges and maintaining focus during both the arduous and rewarding moments of this endeavor.

Lastly, I would like to express my deepest gratitude to the individuals whom I have had the chance to interact with during this thesis who generously shared their insights, experiences, and perspectives.

It is my fervent hope that this Master's thesis contributes to inspiring further research and dialogue and ultimately serves as a catalyst for the growing body of knowledge that is participatory AI, specifically in marginalized communities. May it ignite the necessary discussions, inform policy, and contribute to the development of AI systems that further prioritize fairness, inclusivity, and social impact.

I hope you enjoy the read

- Chakir
June 2023, Delft

Abstract

In today's society, the rapid progression of digitization has led to the automation of various facets of human existence. This transformation has been facilitated by the utilization of algorithms, which are instrumental in driving efficient and effective automated processes. These algorithms have also found widespread adoption in the public sector, where they are employed to streamline and optimize various tasks and operations. The integration of algorithms in the public sector has brought about significant advancements in areas such as predictive policing, social welfare allocation, and healthcare.

However, the use and development of these automated processes were subjected to concerns from the public about privacy, bias, accountability, and transparency. Since these concerns are mainly coming from citizens, their involvement in the process of developing algorithmic systems can potentially be of help.

We explore the potential of participatory AI in marginalized communities as a means of obtaining valuable input from citizens regarding the development of these algorithmic systems employed by the public sector. One Piece of our approach involves hosting discussions in local community centers in marginalized neighborhoods. Our focus is on dilemmas that are relevant to algorithm design and evaluation decisions, and we frame these dilemmas in various ways, including forms that may not directly relate to societal impact, but are understandable for laypeople. Our key findings suggest that involving marginalized citizens can bring valuable perspectives and insights that are otherwise ignored. By incorporating public perspectives into algorithm development, we can promote inclusive decision-making processes and ensure that algorithms align with community values.

Contents

Nomenclature

List of Abbreviations

AI	Artificial Intelligence
DPIA	Data Protection Impact Assessment
FNV	Netherlands Trade Union Confederation
GDPR	General Data Protection Regulation
IAMA	Impact Assessment Human Rights and Algorithms
ICTs	Information and Communication Technologies
IoT	Internet of Things
NJCM	Dutch Section of the International Commission of Jurists
PII	Personally Identifiable Information
SyRi	System Risk Indication

List of Figures

1

Introduction

In the field of mathematics and computer science, algorithms can be thought of as a set of precise instructions or rules, much like a detailed recipe, that guides computers to solve problems or perform specific tasks efficiently [1]. These instructions are akin to a series of logical steps that help computers solve mathematical and computational problems in an organized and systematic manner.

These algorithms come in various forms and complexities. They can range from simple ones to intricate systems that combine multiple 'smaller' algorithms. As a result, algorithms encompass a wide range of possibilities.

Algorithms are growing significantly in ubiquity in modern society and are now widely acknowledged [2, 3]. Decisions and choices previously left to humans are increasingly delegated to algorithms [4]. This paradigm shift has led to algorithms shaping the world we live in today.

In addition to the increasing ubiquity of algorithms, the widespread adoption of surveillance technologies and the growth of the Internet of Things (IoT) has resulted in the creation of a vast interconnected network of data collection devices [5, 6, 7], leading to the availability of massive streams of data.

These massive streams of data combined with increased analytical and technical capabilities play a crucial role in training algorithms and enable researchers, companies, governments, and other public sector actors to resort to data-driven machine learning-based algorithms to tackle complex problems [8, 9, 10]. These algorithms have the capability to leverage this data in order to make predictions or forecasts about various phenomena. By analyzing patterns and relationships within the data, they aim to provide predictions and estimations that can inform decision-making processes.

Applications of these algorithms can be found in a wide range of daily activities. Media consumption in particular is increasingly shaped by automated algorithmic selections, where the selection of online news via search engines or the consumption of music and video entertainment via recommender systems are among prominent examples [2]

Moreover, big data and new approaches to analyzing these data have gained a prominent role in public sector decision-making and public service delivery as well [11, 12, 13].

The public sector is increasingly looking at algorithmic decision-making as a means to reduce costs, optimize bureaucratic processes, improve the quality of decisions, and unleash the power of administrative data, thereby making government performance more efficient

and effective [14, 15, 10, 16]. This pertains to algorithms that are automated and can use machine learning to prioritize citizens in various contexts.

Applications of machine learning tools are used in a variety of domains in the public sector such as criminal justice [17, 18, 19, 20], healthcare [21, 22, 20], and predictive policing systems that predict criminal activity hotspots [4].

However, alongside their widespread adoption, these algorithms pose significant ethical challenges, where citizens are the ones who suffer the consequences. Next to the scale of analysis and complexity of decision-making, the uncertainty and opacity of the work being done by algorithms and its impact are also increasingly problematic, since these algorithms have traditionally required decision-making rules and weights to be individually defined and programmed 'by hand' [4]. It also raises fundamental questions concerning the governance of data, transparency of algorithms, legal and ethical frameworks for automated algorithmic decision-making, and the societal impacts of algorithmic automation itself [3].

If these algorithms are managed poorly by the government, then deployment of these AI tools can have significant negative consequences such as hollowing out the human expertise inside agencies with few compensating gains, widening the public-private technology gap, increasing undesirable opacity in public decision-making, and heightening concerns about arbitrary government action and power [14]. It is thus imperative to exercise caution considering there are already many reported cases in which black box algorithms have caused harm at scale [23].

Because of this, many industry experts, scholars, and activists have already pointed to a range of social, ethical, and legal associated with algorithmic decision-making, including bias and discrimination [24, 25, 10], and lack of transparency and accountability [26, 27, 28, 29, 10] and called for the critical need for a human-oriented explanation by AI systems [23]. Data-driven algorithmic decisions about applicants for jobs, schools or credit may be affected by hidden biases that tend to flag individuals from particular demographic groups as unfavorable for such opportunities [10].

In [30] a call for great caution is exercised for any public sector body considering using complex algorithms as flawed use in this context can lead to harmful consequences for citizens, individually and collectively, and public sector workers.

Real-world scenarios have demonstrated instances of algorithmic bias and discrimination. Two cases that highlight these issues are the Dutch childcare benefits scandal [31] and the automated welfare benefit fraud detection done by the municipality of Rotterdam [32]. Both algorithms used in the cases disproportionally flagged individuals from specific demographic groups exercising bias and discrimination in the process.

In this thesis, we propose a method of engaging with citizens to include them as stakeholders. This might help in the development and deployment of algorithms used by the public sector. At its best, participation leads to individual and collective empowerment as well as social and structural change via the cultivation of new skills, social capital, networks, and self-determination among those who contribute. This has the potential to make a sustained positive impact on the welfare and benefit of communities over time [33, 34, 20]. Technical experts play a crucial role in assisting stakeholders who lack technical knowledge by providing guidance, expertise, and explanations [30] to ensure valuable input from laypeople is effectively incorporated into the development and deployment of these algorithms.

1.1 Thesis objective

The main research objective of this thesis is to identify effective strategies for obtaining valuable input from citizens during the development life cycle of algorithms employed by the public sector. To achieve this, the thesis pursues two primary objectives. Firstly, it aims to raise awareness regarding the concept of citizen input and its significance in contemporary society. Secondly, the thesis conducts an experiment to assess the viability of a specific method as an effective strategy. The ultimate aim is to evaluate the success of the experiment in achieving the overarching research objective. The paper [35] proposes guiding principles for a participatory approach. These principles are designed to help ensure that projects are centered on the needs and values of the communities they impact, particularly those who are most vulnerable or marginalized. Utilizing these principles as inspiration, we have formulated the following research questions to facilitate obtaining our main research objective.

RQ1: What is an effective way of involving marginalized communities?
Involving marginalized communities in the development life cycle of algorithms is crucial as it provides a more comprehensive understanding of existing biases and helps mitigate them. However, it is important to acknowledge that their underrepresentation is not a coincidence but a result of systemic factors. Therefore, our research aims to explore methods to increase the involvement of marginalized communities in algorithm development to address this imbalance and promote inclusive decision-making processes.

RQ2: What are ways to make algorithms more understandable to the public?
Addressing the existing knowledge gaps surrounding algorithms is crucial for harboring transparency and accountability. It is imperative to find methods to eliminate or, at the very least, reduce these gaps. In our experiment, we aim to narrow down this knowledge gap by exploring approaches to make algorithms more understandable and accessible to the public. By doing so, we seek to promote public understanding and engagement with algorithmic systems.

RQ3: How can we effectively incorporate the perspectives of the public?
Incorporating the perspectives of the public into algorithm development is essential for fostering civic participation. However, direct implementation of public perspectives into algorithmic code is not feasible. Therefore, it is crucial to find effective methods to incorporate public perspectives throughout the process. Our research proposes a method to achieve this goal, aiming to establish mechanisms that enable meaningful public participation and ensure that diverse perspectives are considered in algorithm development. By doing so, we seek to enhance the democratic and inclusive nature of decision-making processes in the development of algorithms.

RQ4: How can community engagement be maintained?
Maintaining consistent and valuable perspectives from the public is crucial for improving the long-term development lifecycle of algorithms. To address this, it is essential to explore methods of sustaining community engagement. By answering this question, we can identify strategies and practices that foster ongoing and meaningful participation from the community. This includes establishing effective communication channels, implementing

feedback mechanisms, organizing regular consultations, and promoting transparency and accountability. By maintaining community engagement, we can ensure that the perspectives of the public continue to contribute to the improvement and refinement of algorithms over time.

1.2 Thesis structure

Chapter 2 provides the necessary background information for the study. It elaborates on important overarching definitions, discusses relevant case studies, and describes and evaluates existing structures that are in place. Chapter 3 gives an overview of existing literature regarding algorithms and participatory AI, along with examples of other studies that align with our objective. Chapter 4 discusses key takeaways we took from interviewing field journalists that worked on the case studies, municipality employees, and field experts. Chapter 5 outlines our methodology and experimental setup. Chapter 6 presents some of our experiences in trying to set up and execute our experiment. Chapter 7 presents the obtained results of the experiment. Chapter 8 lays out our findings of the experiments and lists the contributions that were made in the process. Chapter 9 focuses on answering the research questions and explains whether we have successfully achieved our main research objective. Chapter 10 will conclude the thesis with limitations, recommendations, and future work.

2

Literature review

In this section, we will provide an overview of key topics related to this research that can be found in the literature. We will start with the definition of algorithms and the interpretation we used for our study. We then continue with algorithmic governance and its impact. In addition, we will discuss the role of citizens and we will conclude with some insightful case studies.

2.1 Definition of an algorithm

The word 'algorithm' can be traced back to 12th-century Spain when the scripts of the Persian mathematician Muḥammad ibn Mūsā al-Khwārizmī were translated into Latin, which described the methods of addition, subtraction, multiplication, and division using numbers [36, 7]. This translation later 'came to describe any method of systematic or automatic calculation' [37]. Consequently, his last name serves as the origin of the word algorithm as we know it today.

The concept of algorithms is rather abstract, and the word 'algorithm' encompasses a range of meanings across computer science, mathematics, and public discourse [4]. It can be as simple as a cooking recipe or as complex as the backbone of a search engine. As Hill explains, 'we see evidence that any procedure or decision process, however ill-defined, can be called an 'algorithm' in the press and in public discourse' [4].

Computer scientists and mathematicians utilize algorithms to develop efficient solutions for various domains. The instructions they consist of, as highlighted in [1], are a set of logical steps that help computers analyze data, sort information, find patterns, or make decisions, enabling them to solve complex mathematical and computational problems in an organized, precise, and systematic manner.

The definition of algorithms that we employ aligns with the AI definition used by the European High-Level Expert Group on Artificial Intelligence. In their definition, AI refers to 'Systems that exhibit intelligent behavior by analyzing their environment and, with some degree of autonomy, taking action to achieve specific goals' [38]. The Netherlands Scientific Council for Government Policy [39] characterizes AI as a 'system technology', and we do something similar for algorithms, referring to these systems that employ algorithms as algorithmic systems.

There has been a growing interest in enhancing the efficiency and reliability of public sector bodies by implementing algorithmic systems [30]. These algorithms rely on a substantial

amount of data and are therefore often referred to as data-driven algorithms. Data-driven algorithmic decision-making has the potential to improve government efficiency and public service delivery by optimizing bureaucratic processes, providing real-time feedback, and predicting outcomes [16, 10].

2.2 Algorithmic governance

The public sector employs a concept known as algorithmic governance, which encompasses a broad range of sociotechnical practices that regulate and organize society in various ways, including predictive policing, labor management, and content moderation [40]. This phenomenon is evident in areas such as criminal justice [17, 18, 19, 20] and healthcare [21, 22, 20].

As governments increasingly rely on algorithmic systems to make decisions in diverse domains, it has become imperative to establish effective mechanisms for governing these systems. Given the distinctive characteristics of the public sector context, scholars have emphasized the need for additional considerations, such as citizen participation and algorithmic impact assessments [41].

2.3 The public sector has fallen behind

However, at the core of the issue lies the fact that technology often outpaces policy, and this unfortunate reality includes algorithms. Governance mechanisms for algorithms have not kept up with technological advancements [10], such as AI and machine learning.

According to [42], rigid organizational incentives in high-inertia contexts, such as the public sector, hinder rather than support responsible AI work. The difficulties experienced by democratic institutions stem to some extent from an inability to truly embrace the potential of information and communication technologies (ICTs) [43, 44].

Furthermore, algorithmic governance is criticized for its lack of accountability. The functioning of algorithms often defies comprehension by public officials who utilize them, as well as by citizens who are subject to algorithmic decisions and services [45, 13].

Explainability is often used interchangeably with transparency, the latter often having the clearer added meaning of accessibility and interpretability. In nearly every AI ethics research paper, these concepts, along with governance and accountability, are highlighted as highly important and currently lacking, particularly in the public sector where they should be prioritized alongside accuracy [46, 30].

Another issue is that the definition of interpretability and desiderata of what makes a good explanation remain elusive [47, 48], and different researchers use different, often problem- or domain-specific, definitions [23].

2.4 The usual victims

Even the ICTs that have been adopted by the public sector, have been reported to cause harm on a large scale [23]. These harms are also disproportionately experienced by minoritized

populations [49].

The privileged individuals, on the other hand, remain unaffected and are able to maintain their power through a combination of feigning ignorance, discrediting marginalized perspectives, and requiring the non-dominant group to expend their own epistemic resources to combat being silenced and disenfranchised [50, 51, 52].

Minoritized populations face an epistemic burden [52], which is a form of epistemic oppression that arises when these individuals must overcome additional cognitive, emotional, and societal challenges to acquire, produce, and access knowledge compared to more privileged groups. Marginalized communities often encounter systemic barriers and biases that impede their full participation in knowledge creation and dissemination.

2.5 Distrust towards the public sector

Political distrust is also prevalent among these marginalized groups that face discrimination [53]. While the role of distrusting attitudes towards the government has been debated for decades, the dissonance between empirical observations of citizen distrust of politics and the theoretical approaches used to study citizen orientations towards their political system has become more pressing in the past years [53].

Notably, a pattern in the Netherlands emerges where less educated citizens tend to exhibit higher levels of distrust and cynicism towards politics and politicians, whereas well-educated individuals tend to hold more positive views of the government and political institutions [54].

According to [55], the foundation for political distrust among marginalized groups can be traced back to the lack of political representation and the perceived incongruence of interests between these groups and political representatives. Consequently, [55] makes a compelling argument for the need to increase descriptive representation as a means to address and alleviate distrust [53].

In the Netherlands, instances highlighting the emergence of distrust towards the government can be observed in various contexts. One significant example occurred during the recent municipal elections in Rotterdam in 2022, where an unprecedentedly low voter turnout underscored a prevailing sentiment of distrust towards the government [56]. This was especially the case in marginalized neighborhoods where the majority has an ethnic background [57].

Furthermore, the COVID-19 vaccination campaign in the Netherlands serves as another illustrative case, wherein despite the government's active promotion of vaccination, a sense of mistrust arose among certain segments of the population [58]. This distrust was particularly pronounced among individuals with a migration background, as the initial vaccination strategy prioritized individuals without such a background. Professor Agyemang [59], an expert in migration, ethnicity, and health at the University of Medical Care in Amsterdam, has also acknowledged the skepticism experienced by marginalized individuals in this context.

2.6 Citizen participation

Citizens play a crucial role in the functioning of democratic societies. As such, participation and ownership of multiple stakeholders such as citizens have increasingly been promoted [60, 61, 62]. Many scholars and practitioners have explored the benefits of citizen participation, such as increased trust in government, improved decision-making processes, and greater transparency [63].

Data-driven forms of civic participation have become the modern approach for municipalities to engage with citizens, utilizing ubiquitous technology and participatory governance [64]. Researchers, civil society, and regulators increasingly urge greater use of participatory methods to mitigate sociotechnical risks [65, 20].

Participation leads to individual and collective empowerment as well as social and structural change via the cultivation of new skills, social capital, networks, and self-determination among those who contribute. This has the potential to make a sustained positive impact on the welfare and benefit of communities over time [33, 34, 20].

Co-design is one of the approaches for creatively engaging citizens and stakeholders to develop new solutions to complex problems [66, 13] and research has highlighted their potential benefits along with co-creation [13].

These benefits include cost-effectiveness, access to diverse perspectives, fostering stronger relationships with citizens, moving away from the bureaucracy in traditional systems, reinforcing a sense of belonging and citizenship, and overall increasing capacity to respond to citizens' needs [67].

An example is the study conducted by [13] that demonstrated co-design playing a crucial role in creating an algorithm that was understandable to its users. By being part of the design process from the onset, it was clear to those who would be using the tool what data sources the algorithm uses as input and what indicator it uses to make the necessary estimations.

Another approach to citizen participation is through participatory decision-making processes, which recognize and involve citizens as active partners in public sector management, enabling them to contribute their expertise, experience, and knowledge.

2.7 Inclusive stakeholder engagement

The automation of human decision-making is often defended by claiming that algorithms are unbiased [68, 69]. However, algorithms inevitably exhibit biases, whether conscious or unconscious, leading to subtle influences on people's perceptions and actions, and as a result discrimination. Fundamentally, an algorithm's design and functionality reflect the values and beliefs of its creator and its intended purposes [4].

The paper [24] argues that discrimination may be an artifact of the data collection and analysis process itself. The collected data may suffer from preferential sampling, resulting in a biased dataset. Algorithmic models may use data sources and indicators that are ill-suited for the purposes of public governance which can lead to incorrect and discriminative decisions [10].

Additionally, societal biases can infiltrate algorithms through the data collected, reflecting existing biases within society [70, 3]. Data sets are snapshots of the past, encompassing any

previous mistakes, statistical biases (i.e., skewed data representation rather than prejudice), past environmental conditions, or outdated policies. These biases become embedded in the data and can be perpetuated by the algorithm when applied in the present [30].

More specifically, even with the best intentions, data-driven algorithmic decision-making can lead to discriminatory practices and outcomes: algorithmic decision procedures can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society [71].

For instance, data-driven algorithmic decisions concerning job applicants, school admissions, or credit evaluations may be influenced by hidden biases that disproportionately disadvantage certain demographic groups. These outcomes can create a self-reinforcing cycle, as reduced access to credit, employment, and education worsens individuals' circumstances and hinders their future applications [10].

An illustrative example is the depiction of African-Americans as more likely to commit violent crimes compared to whites in the COMPAS software, which uses algorithms to assess potential recidivism risk. The results inform future decisions for the software to revise the algorithm to attain a certain rate of accuracy. This cause-and-effect sequence illustrates how the algorithm can increasingly perform better at the cost of bias towards African-Americans, ultimately resulting in a feedback loop [72].

2.8 Case studies in participation

Several case studies exist that explore various instances of citizen participation in the design and development of algorithms.

The case study presented in [13] explores the co-designing process in the algorithmic management of supplies for refugee camps. The key findings were that co-designing supported the responsibility and accountability of the algorithm by making the estimations transparent and explicable to its users. The inclusion of knowledge from the field led to changes in the algorithm at different stages of development. Co-design also contributed to preventing the reinforcement of biases and supported the responsible selection of big data sources. However, it does have to be noted that it remains unclear whether co-design eventually led to better servicing of refugee camps.

In [73], a participatory algorithmic governance framework to shift work schedules as a case study has been proposed. They worked with 28 shift workers and scheduling who used their web tool to build their well-being models and shared their experience through interviews. The results of this case study suggest that incorporating worker well-being into algorithmic management can optimize workplaces for workers. The authors argue that it is critical to computationally model worker well-being and directly incorporate it into algorithmic workplace design. The study also illuminates the opportunities and challenges in defining worker well-being for algorithmic management.

The study described in [74] used citizens' juries as a form of deliberative democracy to elicit informed judgment from a representative sample of the general public around policy questions. Per scenario, jurors voted for their preferred system; votes were analyzed descriptively. Qualitative data on considerations behind their preferences included transcribed audio recordings of plenary sessions, observational field notes, outputs from small group work, and free-text comments accompanying jurors' votes; qualitative data were analyzed

thematically by scenario, per, and across AI systems. The study concludes that citizens may value the explainability of AI systems in healthcare less than in non-healthcare domains and less than often assumed by professionals, especially when weighed against system accuracy. The public should therefore be actively consulted when developing policy on AI explainability.

These case studies provide valuable insights into the practical applications of algorithms across different domains, emphasizing the importance of considering various factors such as goal-setting, data quality, inclusive design processes, and public engagement in algorithmic decision-making.

3

Background

In this section, we will be discussing additional background knowledge essential for understanding the motivation behind this research derived from grey literature. We will be providing more information on the public sector, as well as elaborating on the case studies in the Netherlands that motivated this thesis. Finally, we will discuss existing structures that are already in place as a result of these case studies.

3.1 Public sector

As mentioned, the public sector using algorithmic systems provides an increase in efficiency in decision-making and policy management which betters the services to its citizens.

An area where algorithms have been implemented and deployed in the Netherlands is crime prediction, specifically, fraud detection [75]. Through analyzing data, the data-driven algorithm would be able to predict patterns and anomalies that would suggest fraudulent activity. This enables the proper authorities to take appropriate action.

While these algorithms can significantly improve decision-making and service delivery, they can also be exposed to potential risks associated with their use. The data used may contain biases leading to discriminatory outcomes. The public sector is therefore responsible for mitigating these risks to ensure that the ethics of the used algorithmic systems are in order. The responsibility also pertains to privacy, accountability, and transparency.

3.2 Case studies fraud detection

We will discuss four case studies that highlight the potential pitfalls of relying solely on automated decision-making processes without sufficient human oversight. They underscore the urgent need to examine the implications of algorithmic decision-making in highly sensitive domains.

The first case study revolves around the childcare benefit scandal referred to as 'toeslagenaffaire' in the Netherlands. It is an infamous scandal that exposed the devastating effects of an algorithmic system that was deployed to detect fraud in childcare benefits and served as a trigger for media attention regarding other algorithms employed by the public sector. The second and third case studies focus on algorithms employed by municipalities that detect welfare fraud. Lastly, we turn our attention to System Risk Indicator (SyRi), a

controversial algorithmic tool developed in the Netherlands for identifying potential welfare benefit fraud.

3.2.1. Childcare benefit scandal

The childcare benefit scandal ('toeslagenaffaire') refers to a scandal that took place in the Netherlands and involved its Tax and Customs Administration and government [31, 76]. The Tax Administration wrongly accused thousands of parents of fraud in claiming childcare benefits. Unfortunately, these false accusations were not rectified in a timely manner letting parents face many hardships. These hardships included the burden of repaying substantial amounts of money, the requirement to provide evidence for every transaction, and enduring significant stress.

The scandal was big enough for the entire Dutch government to resign on January the 15th 2021 [77]. The government's failure to address the issue adequately despite prior awareness prompted this drastic measure. This resulted in a public outrage after everything was made public by investigative journalists and politicians.

These journalists and politicians [78] discovered that the primary cause of the wrongful accusations stemmed from a flawed algorithm. The government introduced this algorithm in 2013 to automate the detection of fraudulent applications for childcare benefits.

In short, the algorithm used a risk profile that was based on various characteristics of the applicants. These characteristics would include their nationality, ethnicity, and postal code.

Such profiling led to a disproportionate number of individuals with a migration background being wrongly accused of fraudulent activities. This hints at discrimination and racism.

Not only was the discriminatory factor an issue, but the lack of transparency in the decision-making process also became very evident. Thus the use of algorithms in the childcare benefit scandal has raised many questions about the ethics of using algorithmic systems in essential decision-making processes. Particularly areas that have a significant impact on people's lives.

3.2.2. Municipal Welfare algorithms

Two notable algorithms emerged that were employed by municipalities with the aim of effectively detecting fraudulent welfare applicants. However, as subsequent investigations revealed, these algorithms were ultimately exposed as flawed and contained biased and discriminatory behavior.

Rotterdam's social welfare algorithm

The social welfare fraud detection algorithm used by municipalities in the Netherlands including Rotterdam is another case where things took a turn for the worst when using these algorithmic systems. Even though this algorithm was put to use by multiple municipalities, investigative efforts by journalists of Lighthouse Reports [79], Argos [32], Vers Beton [80], and Follow the Money [81], brought the municipality of Rotterdam into the spotlight.

The algorithm used was designated to identify potential cases of social welfare fraud. This was done by analyzing various data sources, such as income, tax records, and property

ownership records. The algorithm was supposed to be a cost-effective method to reduce spending money on fraudulent benefactors of welfare in the municipality. However, the algorithm was found to be discriminatory against certain people and led to false accusations of fraud.

Investigative work showed that, for example, young mothers and people who are not proficient in the Dutch language were indicated as much more likely to be fraudulent benefactors of welfare.

The discriminatory impact of the algorithm led, just like in the case of the childcare benefit scandal, to disastrous consequences for those falsely accused. Economic problems stemmed from these accusations and invasive time-consuming investigations into the lives of the people also worsened their situation.

The impact of the deployment of the algorithm led to widespread criticism and in turn, led to the eventual shutdown of the project for all municipalities. Rotterdam, together with the other municipalities were left to review the impact of these algorithms on human rights.

This case again highlights the potential negative impact of algorithmic systems on human rights, especially in delicate situations such as citizens who require social welfare.

Fraud scorecard

Another algorithm called the 'fraud scorecard' [59, 82] was developed by the government in 2003 and used by 158 municipalities to predict welfare fraud. It was supposed to have been abolished in 2020. However, up until 2022, four municipalities used a derivative version and eventually also quit using the algorithm after being questioned by journalists.

The algorithm assigned points based on factors such as profession, living situation, education level, gender, and neighborhood. For example, caravan residents received 700 points while homeowners received 0 points. If a person's score reached 950 points, the computer would indicate they are fraudulent.

Since the algorithm was developed in 2003, welfare applications were weighed based on risk profiles established twenty years ago. The municipalities did not know why citizens were flagged as suspicious because they did not have access to the underlying calculations. This is in violation of privacy law (GDPR).

3.2.3. SyRi

SyRi (Systeem Risico Indicatie) is a risk indication system used in the Netherlands. It is an automated data analysis tool developed by the Dutch government to assess the risk of fraud or irregularities in social security benefits and tax compliance. SyRi analyzes various data sources, such as income, assets, housing, and personal information, to identify potential discrepancies or anomalies that may indicate fraudulent activity.

However, several civil society interest groups, including the Dutch Section of the International Commission of Jurists (NJCM) and two private individuals, instituted proceedings against the State of the Netherlands. The Netherlands Trade Union Confederation (FNV) joined as a party in the claimants' proceedings [83].

The Hague District Court ruled that the legislation regulating the use of SyRi violates higher law. According to the court, there was no balance between the societal interest

and the infringement on people's private lives. As a result, the regulations underlying the instrument SyRI have been deemed to be in violation of higher law, and the court declared the regulations to be non-binding.

At first, the State disagreed arguing that SyRi legislation contained sufficient safeguards, but later decided against and discontinued the program indefinitely.

3.2.4. Patterns of mistakes

When analyzing the presented case studies, several patterns of what went wrong with these algorithmic systems emerge. Discrimination and bias stood out prominently in the childcare benefit scandal and the municipality algorithms, where the algorithms used characteristics such as nationality, and ethnicity which led to disproportionate false accusations against individuals with a migration background.

Transparency, or the lack thereof, formed another prevalent pattern in these case studies. The childcare benefit scandal exposed the algorithm's inadequate transparency, leaving individuals unaware of how decisions were made. Similarly, the municipality algorithms also suffered from a lack of transparency. Even the municipalities did not have access to information in the fraud scorecard case.

All of them also had a lack of ethical considerations as these algorithms were employed in critical areas such as childcare benefits and social welfare, which have a significant impact on people's lives. Their main priority was efficiency and cost savings over protecting the livelihood of their citizens.

3.2.5. Potential solutions

To address these mentioned issues, several solutions can be implemented. Enhancing algorithmic transparency by disclosing criteria and data sources, and providing clear explanations promotes needed accountability and trust.

The implementation of citizen panels or another participatory process also emerges as one of the solutions. Involving diverse groups of citizens in the design, development, and oversight of algorithms can bring valuable perspectives and insights. This helps to prevent biases from occurring and ensures algorithms align with community values.

Promoting a change in the environment of developers and policymakers can potentially be an effective solution. By fostering diversity and ensuring a representation of society, we can create an environment that incorporates a broader range of perspectives. This inclusivity can bring valuable insights and considerations to the table, enhancing the decision-making process.

Finally, better collaboration and communication between already involved stakeholders, such as developers, government agencies, civil society organizations, and academic institutions may also prove to be an effective medicine to the sickness that is discrimination.

3.3 Algorithm register

One approach to promoting transparency and accountability in the public sector is through the implementation of algorithm registers. This concept has been introduced in several

municipalities in the Netherlands.

Algorithm registers serve as documentation systems that keep records of the algorithms used by these municipalities. They typically include essential information about the algorithms, such as their names, descriptions, and the reasons behind their deployment. By documenting and making this information publicly available, municipalities aim to foster trust and understanding among citizens and stakeholders.

In the context of our study, we will specifically examine the algorithms employed by the four largest municipalities in the Netherlands: Amsterdam, Rotterdam, The Hague, and Utrecht. We will delve into the details of these respective algorithms to gain a deeper understanding of their functions and implications.

3.3.1. Amsterdam

Amsterdam adopts a proactive approach to promoting transparency and accountability by publicly disclosing their algorithms on their official website. This practice aims to enhance the understanding and scrutiny of their algorithmic systems. In Figure 3.1, an example is displayed of detailed information that is provided by Amsterdam about a specific algorithm, which in this case is about scanning license plates of cars that have been illegally parked, showcasing Amsterdam's commitment to openness.

Figure 3.1: Screenshot of detailed information on an algorithm in Amsterdam's algorithm register.[1]

Each algorithm is accompanied by a comprehensive description that delves into the underlying reasoning behind its design and implementation. Amsterdam articulates the anticipated benefits and outcomes associated with the utilization of these algorithms. Technical aspects, such as the specific datasets employed and the architecture of the model used, are also provided, further enhancing the transparency of their algorithmic decision-making processes.

[1] Accessed via https://algoritmeregister.amsterdam.nl/ on 3-6-2023.

Notably, Amsterdam recognizes the significance of incorporating ethical, religious, and political considerations to ensure the prevention of discrimination and the promotion of fairness. By addressing these dimensions, Amsterdam strives to develop algorithmic systems that align with societal values and norms.

Moreover, Amsterdam emphasizes the presence of human oversight throughout the utilization of these algorithms, acknowledging the critical role of human judgment and intervention in decision-making processes. This human oversight serves as a safeguard against potential biases or unintended consequences that may arise from the automated nature of the algorithms.

Amsterdam acknowledges the risks associated with the use and management of algorithmic systems, underscoring the need for effective governance and oversight mechanisms. By openly discussing these risks, Amsterdam demonstrates their commitment to responsible algorithmic practices and the continuous improvement of their algorithms' performance and impact.

3.3.2. Rotterdam

Similarly to Amsterdam, Rotterdam discloses algorithms on its official website. An example of these algorithms is shown in Figure 3.2. The figure shows how Rotterdam displays detailed information about the algorithm. In this case, it is about recognizing licensing plates of cars that have been illegally parked.

Figure 3.2: Screenshot of detailed information about an algorithm in Rotterdam's algorithm register on their website[2]

Their register consists of four main sections: description, verifiability, governance, and risk mitigation.

The description, much like Amsterdam, provides an overview of the reasoning behind the algorithm's design and implementation. The verifiability section focuses on the auditability and transparency aspects of the algorithm. It explores the concept of explainability, the

[2]Accessed via https://algoritmeregister.rotterdam.nl/p/Onzealgoritmes on 3-6-2023.

specific data fields involved in personally identifiable information (PII), and the accessibility of the algorithm's casuistic foundation.

The governance section expands upon the information presented in the description section, emphasizing the importance of effective administration and oversight. It also outlines an appeal process that citizens can pursue if they wish to contest the algorithm's outcomes or decisions.

Lastly, the risk mitigation section delves into the measures taken to manage potential risks associated with the algorithm. It discusses the presence of human oversight, the legal basis on which the algorithm operates, its risk classification, and the performance of a Data Protection Impact Assessment (DPIA) to identify and address privacy and ethical considerations.

3.3.3. The Hague

The algorithm register of The Hague is in the form of a document. Part of this is shown in Figure 3.3. The document describes algorithms that have been developed. Each row represents an algorithm they have developed, and each column gives some metadata about the algorithm.

The algorithm register of The Hague municipality is presented in the form of a document, and a portion of it is displayed in Figure 3.3. Each row in the document represents an individual algorithm, while each column contains metadata associated with the algorithm.

Naam algoritme	Taal	Laatst bijgewerkt	Naam algoritme	Organisatie	Dienstonderdeel	Korte omschrijving	Soort algoritme	Website	Status	Besluitvorming
	Nederlands	Datum laatste revisie van dit document	De gangbare naam gebruikt om het algoritme aan te duiden.	Gemeente Den Haag	Dienst/afdeling	Algemene, korte beschrijving van het algoritme, max 150 chars	Op regels gebaseerd of zelf lerend	De url referentie naar de openbare publicatie van de code van het algoritme indien van toepassing.	ideenfase/ in ontwikkeling/ in productie/ gestopt	Bij welk besluitvormingsproces in de organisatie is het algoritme betrokken. Is er sprake van geautomatiseerde besluiten. Indien mogelijk verwijzen naar concrete besluitvorming of beleid, zoals ontsloten via publieke bronnen als raadsinformatiesystemen
Wmo-voorspelmodel	Nederlands	19-12-2022	Wmo-voorspelmodel	Gemeente Den Haag	Onderwijs, Cultuur en Welzijn	Het Wmo-voorspelmodel geeft een voorspelling van het aantal unieke gebruikers en de kosten van de Wmo met een voorspelhorizon van zes jaar. Voorspellingen worden gedaan op wijkniveau voor zowel de Wmo-totaal als voor deelproducten van de Wmo (Hulp bij het Huishouden, Ondersteuning thuis en Hulpmiddelen en Diensten).	Op regels gebaseerd		Actief	Met de inzichten uit het Wmo-voorspelmodel wordt beleid en uitvoering ondersteund bij beantwoorden van tactische en strategische vragen. Dit zijn vragen en onderwerpen die op de langere termijn spelen (bijvoorbeeld 5 á 10 jaar). Het model is niet bedoeld voor het beantwoorden van operationele vragen (onderwerpen die op korte termijn spelen) en er is geen sprake van geautomatiseerde besluiten.
Scanauto kentekenherkenning	Nederlands	28-10-2021	Scannen voor parkeerhandhaving en handhaving milieuzones	Gemeente Den Haag	Stadsbeheer	Het doel van het algoritme is om kentekens automatisch te scannen ten behoeve van parkeerhandhaving en handhaving milieuzones.	Op regels gebaseerd		Actief	Met het algoritme wordt het proces ondersteund om te bepalen of een auto fout geparkeerd is (Fiscale Parkeerhandhaving) en om te bepalen of een auto/vrachtauto ten onrechte in een bepaald gebied rijdt (Handhaving Milieuzones).

Figure 3.3: Screenshot of part of the algorithm register of The Hague in XLSX format[3]

The metadata associated with each algorithm provides crucial information about its characteristics and usage. It includes the algorithm's name ('Naam algoritme'), the organization or municipality employing the algorithm ('Organisatie'), the specific service element to which the algorithm applies ('Dienst onderdeel'), and a brief description of its functionality ('Korte omschrijving'). The algorithm register also includes maintenance details such as the version number and the date of the most recent update ('Laatst

[3]Downloaded from The Hague Data Platform via https://denhaag.dataplatform.nl/#/data/f58f2b0e-8d93-480c-b6f7-da95ed7bbe18. Accessed on 3-6-2023.

bijgewerkt'), ensuring that the information provided remains current and accurate.

To enhance clarity, the algorithm register specifies the type of algorithm used, allowing for a distinction between rule-based and self-learning algorithms ('Soort algoritme'). The register also includes information about the methods and models employed in the design and implementation of the algorithm. In instances where more extensive information is available, a link to a website containing additional details ('Website') may be provided. This website can potentially offer access to additional resources such as the algorithm's source code, offering interested individuals a deeper understanding of its inner workings.

The algorithm register also indicates the operational status of each algorithm, specifying whether it is currently in use ('Status'). Moreover, the register provides a comprehensive description of the policy framework associated with each algorithm. This includes details regarding its legal basis, function, and overarching goal, providing essential context to understand the purpose and impact of the algorithm.

In order to assess the potential effects and address risks associated with the algorithm, a concise analysis of its impact and any monitoring risks is provided. Moreover, the algorithm register outlines alternative approaches that have been explored, if applicable, demonstrating the careful consideration given to different solutions.

If the algorithm deals with personal information, the specific personal data used and a description of its nature are outlined. Additionally, the responsible party, their obligation to provide information to the public, and the available appeals process are detailed to ensure transparency and accountability.

3.3.4. Utrecht

In a manner similar to The Hague, Utrecht employs a documentation approach to outline its algorithms, as depicted in Figure 3.4. Utrecht specifies that the document excludes algorithms categorized as 'small' to ensure a clear overview and manageable structure [84], although the definition of a 'big' algorithm is not explicitly provided.

1. Naam van het proces	2. Omschrijving van het proces	3. Wat is de rol van het algoritme of voorspellend model binnen het proces?	4. Geeft het algoritme of voorspellend model informatie of neemt het zelfstandig een besluit?	5. Is een Privacy Impact Assessment uitgevoerd?	6. Is een Uthiek Assessment uitgevoerd?
Dynamisch inzamelen afval	Het voorspellen van het juiste moment waarop afvalcontainers geleegd kunnen worden en daar de route en het schema van de vuilniswagen op inrichten	Het algoritme combineert informatie van de sensoren mbt de vulgraad in de ondergrondse afvalcontainers tot informatie voor de meest effectieve rijroute en schema van de vuilniswagens	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste woord, maar een medewerker.	Ja	Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments uitvoerden
Slim vegen	Het voorspellen van de meest nuttige rijroutes van veegwagens	Het algoritme combineert historische data mbt rijroutes van veegwagens tot informatie voor de meest effectieve rijroute en schema van de veegwagens	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste woord, maar een medewerker.	Ja	Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments uitvoerden
Vervanging lichtmasten	Het voorspellen van het juiste moment waarop lichtmasten aan vervanging toe zijn	Het algoritme combineert criteria aan een bestand van lichtmasten tot informatie over welke lichtmasten in aanmerking komen voor vervanging	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste woord, maar een medewerker.	Nee, in overleg met de DISO en de FG is bepaald dat een Privacy Impact Assessment niet nodig is	Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments uitvoerden
Stedelijk water	Het voorspellen van de hoogte van het stedelijk water en de riolering	Het algoritme combineert historische data van de sensoren mbt waterstanden van het stedelijk water tot informatie over de waterstanden in de nabije toekomst	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste woord, maar een medewerker.	Nee, in overleg met de DISO en de FG is bepaald dat een Privacy Impact Assessment niet nodig is	Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments uitvoerden
Onderhoud wagenpark	Het voorspellen van het juiste moment waarop onderhoud nodig is bij het wagenpark	Het algoritme combineert criteria aan een bestand van wagenparken en geeft informatie over welke wagenparken wanneer in aanmerking komen voor onderhoud	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste woord, maar een medewerker.	Ja	Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments uitvoerden
	Het geautomatiseerd herkennen van	Het algoritme combineert historische data mbt kenmerken van verschillende typen straatafval tot informatie over welke type	Het algoritme geeft informatie en dit stelt een medewerker in staat om, mede op basis hiervan, het juiste besluit te nemen. Het algoritme heeft dus niet het laatste		Nee, omdat we bij aanvang van deze toepassing nog geen Uthiek Assessments

Ruimte | Stadsbedrijven | VTH | Vastgoed | MO | VG | W&I | Wijken | Veiligheid | PBZ | BSN | RO

Figure 3.4: Screenshot of part of the algorithm register of Utrecht in XLSX format. The different fields are shown in the bottom tabs. [4]

The document's contents, akin to The Hague, encompass metadata summaries of their algorithms. The bottom tabs indicate the differentiation between various predefined fields. However, it is worth noting that Utrecht's document contains substantially less information compared to that of The Hague. Utrecht's approach encompasses limited details, such as the process name ('1. Naam van het proces'), a process description ('2. Omschrijving van het proces'), the algorithm's or predictive model's role within the process ('3. Wat is de rol van het algoritme of voorspellend model binnen het proces?'), and whether the algorithm or predictive model provides information/predictions or autonomously makes decisions ('4. Geeft het algoritme of voorspellend model informatie of neemt het zelfstandig een besluit?').

Additionally, the document indicates whether a 'Privacy Impact Assessment' has been conducted ('5. Is een Privacy Impact Assessment uitgevoerd?'), which evaluates the impact of the process on privacy, as implied by its name [85]. Lastly, the document specifies whether an 'Uthiek Assessment' has been conducted ('6. Is een Uthiek Assessment uitgevoerd?'), which refers to Utrecht's ethical values model encompassing principles such as privacy, justice, safety, health, and more [86].

3.3.5. Criticism

The algorithms used by these municipalities are also recorded in a national algorithm register [87], along with those of other governing bodies. Despite this, there are still differences in approach among the municipalities. This pertains to both the manner of presentation and the extent of information provided. For instance, Rotterdam has only made 4 algorithms available, whereas Utrecht has disclosed 38, excluding the smaller algorithms mentioned in section 3.3.4. There seems to be a lack of unity, which is somewhat unexpected.

Furthermore, the descriptions of the algorithms that are provided often contain empty fields, lacking crucial information. The provided descriptions can also be vague. Specific criteria determining the importance of an algorithm for inclusion in the algorithm register are not defined. There is significant room for improvement in establishing a consistent approach that encompasses all necessary aspects, such as data ethics and verifiability [88].

Nevertheless, these algorithm registers represent a positive step toward promoting transparency and accountability.

3.4 IAMA

In response to various incidents, including the child benefit scandal, the Dutch government adopted the IAMA [89, 90], which was introduced and developed by researchers of Utrecht University on behalf of the Dutch Ministry of the Interior and Kingdom Relations [91]. IAMA [92] stands for 'Impact Assessment Mensenrechten en Algoritmes' which translates to 'Human Rights Impact Assessment and Algorithms'. It is a document that refers to a process of evaluating the potential impacts of algorithms on human rights and developing strategies to mitigate any negative effects. It was designed as a preventive tool to avoid any similar event from occurring ever again.

[4]Downloaded from https://data.utrecht.nl/dataset/algoritmeregister-utrecht. Accessed on 3-6-2023.

The goal of this assessment is to ensure that algorithms and automated decision-making systems do not violate human rights, such as the right to privacy, non-discrimination, and freedom of expression. The IAMA typically involves a thorough analysis of the design, implementation, and use of algorithms, taking into account their potential impacts on individuals and society as a whole.

3.4.1. Content

The IAMA framework can be divided into three phases. The first phase is titled 'preparation' where the 'why' is answered. The second phase titled 'input and throughput' consists of two sub-phases; 'input', which discusses data, and 'throughput', which discusses throughput. The overall question answered is the 'what'. Finally, the third phase has how as its central question. In addition to the three phases an additional part called 'human rights' is explained that discusses broader questions that overarch the three phases, including fundamental rights.

Phase 1 (Why)
In this initial phase, the primary objective is to clearly define the purpose for the development of the algorithmic system. This involves identifying the societal problem that the algorithm aims to address. Additionally, an assessment of the potential benefit and risks associated with using the algorithmic system.

Phase 2 (What)
2A (Input)
The focus of this sub-phase is to identify and evaluate the data used as input for the algorithmic system. An assessment is conducted to determine the quality and reliability of the data used. Special attention is given to identifying potential biases and inaccuracies that could negatively affect the algorithmic system and ensure fairness.

2B (Throughput)
The second part of phase 2 involves an in-depth analysis of the algorithmic system itself. This entails examining the training methods and models employed. Identifying and subsequently resolving any potential biases and critical issues within the algorithmic that may arise during deployment are crucial components of this sub-phase.

Phase 3 (How)
This phase involves evaluating the impact of the algorithmic system on citizens by assessing whether the outcomes are positive or negative. All human rights should be upheld and any violations that may arise should be detected. Furthermore, the effectiveness of the algorithmic system in addressing the societal problem is also assessed.

Human Rights
Finally, the IAMA contains an additional step that involves developing and implementing strategies to mitigate any potential negative impacts that may have risen, especially human rights violations. This step aims to ensure the preservation of fundamental rights in the context of algorithmic systems used by the public sector.

3.4.2. Involved people

The IAMA framework involves the active participation of various stakeholders to ensure the responsible and ethical development and implementation of algorithms. These stakeholders include:

- Stakeholder Group (Belangengroep): Represents the interests of those affected by the algorithm, ensuring their perspectives are considered.
- Boarc (Bestuur): Responsible for overseeing the project, and providing guidance and accountability.
- Citizen Panel (Burgerpanel): Provides valuable input and feedback from the citizen's perspective, ensuring their voices are heard in the decision-making process.
- Chief Information Security Officer/Chief Information Officer (CISO/CIO): Ensures information security and safeguards data privacy throughout the algorithmic process.
- Data Scientist (Databeheerder of Bronhouder): Analyzes and interprets data, applying expertise in data management and analysis.
- Data Manager/Data Source Holder: Responsible for the management and integrity of data used in the algorithm.
- Domain Expert (Domeinexpert): Brings domain-specific knowledge and expertise to ensure the algorithm aligns with the specific application area.
- Data Protection Officer (DPO): Ensures compliance with data privacy regulations and safeguards individuals' rights regarding their personal data.
- HR Employee (HR-medewerker): Manages personnel involved in the project, addressing HR-related matters.
- Lawyer (Jurist): Provides legal advice and ensures the algorithm and its implementation comply with relevant laws and regulations.
- Algorithm Developers (Ontwikkelaar algoritme): Responsible for designing, coding, and testing the algorithm.
- Client/Organization Requesting the Algorithm (Opdrachtgever): The entity requesting the algorithm and benefiting from its outcomes.
- Other Members of the Project Team (Overige leden projectteam): Collaborators and stakeholders with specific roles and responsibilities within the project.
- Project Leader (Projectleider): Leads and coordinates the project, ensuring its successful execution.
- Strategic Advisor on Ethics (Strategisch adviseur ethiek): Provides guidance on ethical considerations and ensures the algorithm aligns with ethical principles.

The involvement of these diverse stakeholders ensures that the algorithm is developed and implemented in a responsible and ethical manner while considering the interests of all parties involved. However, the high number of stakeholders involved might impede the process of progressing through the IAMA accordingly.

In this research, particular emphasis will be placed on the role and significance of the citizen panel, which plays a crucial part in ensuring the algorithm aligns with societal values and addresses the needs and concerns of citizens.

3.4.3. Limitations of described citizen panel use

It is clear that the IAMA recognizes the significance of citizen panels and their role in assessing the impact of algorithms on human rights as well as detecting otherwise unnoticed biases that transpire. Including citizens in the process promotes transparency and accountability, however, the current implementation of citizen panels described has its limitations.

According to the document, citizen panels are primarily involved in the initial stage ('Why' phase), where they provide input on the intended behavior of the algorithmic system. This means that biases and potential human rights violations could still go unnoticed in later stages. To ensure this does not happen, we advocate for a more consistent involvement throughout the process.

Furthermore, the document states that these citizen panels are suggested to be involved in the discussion and also states that utilizing them is an option next to involving a representative of an interest group. This means citizen panels don't even need to be used, so algorithms can be developed without any involvement from the citizens.

Thus, it remains important to consider the limitations of the citizen panel as described in the IAMA. While involving citizens can provide valuable insights, they should consistently be involved throughout the entire process of development and deployment. Additionally, robust mechanisms must be set in place to ensure human rights are not negatively impacted. Properly promoting transparency and accountability is essential in the effective utilization of citizen panels.

3.5 Challenges in citizen panels

Next to the limitations described around the involvement of citizen panels in the IAMA, there are also other weaknesses associated with the use of a citizen panel, specifically within the citizen panel itself.

One of the challenges is the subjective nature of fairness and its various definitions [93, 94, 95]. While the objective is to have a fair discussion among the participants, the notion of fairness remains subjective and varies depending on people's social preferences. Even researchers are still unable to formulate a precise definition. If the participants have different views on fairness, they might hinder the effectiveness of the citizen panel. Additionally, if the participants do not feel that their definition of fairness is being heard or respected, they may disengage from the discussion, which could lead to further polarisation and a lack of progress. It is therefore important to acknowledge and manage these subjective perspectives to maintain a fair and balanced dialogue.

Recruiting a diverse and representative sample of participants can also be challenging [96]. To mitigate selection bias, it is important to use a variety of recruitment methods and ensure that efforts are made to reach a diverse pool of participants. However, determining the level of diversity can be challenging. In terms of diversity, various factors should be considered, including but not limited to, ethnicity, religion, and gender [97, 98]. An example would be the American jury selection during court where jury members are randomly selected from a diverse pool with enough representation of minorities [99], but can at the same time result in no diversity at all if you are 'unlucky'.

Another concern is the lack of expertise present among the participants of the citizen panel [100, 98]. They may be asked to provide input regarding complex issues or very technical matters that require a certain level of expertise or some specific knowledge. However, these participants are to be considered laypeople and are not expected to have the necessary background to fully understand these issues. This restricts the quality and quantity of their contributions. They are also more subjected to biased framing favoring the use of the system [98]. This can potentially lead to uninformed decisions they might actually not agree with.

Concentrating on the acceptance of citizen panels and social acceptance of these algorithms threaten to obscure moral issues. It is therefore helpful to combine both social acceptance and ethical acceptability as they are complementary [101].

Time constraints [102] can also impact the effectiveness of a citizen panel. If participants are not given enough time to fully understand, let alone consider and form opinions on the topic at hand, their feedback may be rushed. This can lead to inconsistent responses that are not an accurate representation of their views.

Groupthink [103] is another potential pitfall of a citizen panel. Groupthink is the phenomenon when a group of people is more concerned with reaching a consensus rather than their perspectives that differentiate. This can result in suboptimal or sometimes the poorest decisions and can limit the range of opinions and ideas that are considered in the process.

Another issue is limited participation or (unexpected) changes in participants we found in our study. This reduces the effectiveness of the citizen panel, since a smaller panel or a panel where participants are replaced mid-process may provide feedback that is not representative enough. Additionally, if the panel is subject to too many changes, it will make it more difficult to build on previous discussions.

Ensuring equal participation and contribution among the participants is crucial [62]. It is important to note that certain individuals may feel uncomfortable speaking up during discussions. Other participants, on the other hand, may dominate the conversation with their opinion, hindering the contribution of others in the process. This highlights the need for an environment where every participant is comfortable enough in expressing their opinions and encouraged to listen to others as well.

3.5.1. Community engagement

For active community engagement rather than simply community outreach, an incentive is an important aspect to consider [104]. Intrinsic motivation is more powerful than extrinsic motivation. But creating intrinsic motivation and incentives can be rather difficult. Individuals who believe their contribution is insignificant or ignored are less likely to engage in participatory processes. Additionally, individuals who feel disconnected from the decision-making process or believe that their input is not valued may also be less motivated to participate.

Therefore, it is important to ensure that participants feel their involvement is meaningful and that their input is valued to encourage continued engagement in the decision-making process [105]. This brings out the intrinsic motivation to keep the participants incentivized and maintain community engagement.

Overall, by intentionally seeking out diverse perspectives and creating a safe and inclusive

space for discussion, we can ensure that the citizen panel is representative of the broader community and can provide valuable insights and recommendations for the public sector.

4

Interviews

As part of our exploration of citizen participation and citizen panels, we conducted interviews with individuals holding positions that we believed could provide valuable insights.

Through recommendations and our network, we connected with individuals who possessed expertise relevant to our research. These interviews were conducted to gather diverse perspectives, including those on citizen panels. This section presents an overview of the interviews, highlighting the key findings. Throughout the interviews, we identified three primary categories of interviewees: municipalities, journalists, and civil society organizations.

Given the sensitive nature of our topic, we have taken the decision to anonymize the individuals interviewed. In Table 4.1, you can find the encoded identifiers of the interviewees, along with their affiliations and respective roles within those organizations. This approach ensures the protection of their privacy.

Role	Encoding	Affiliation
Alderman	M1	Municipality
Council member	M2	Municipality
Head of Algorithm Project	M3	Municipality
Head of Strategy and Support	M4	Municipality
Digital Strategy Consultant	M5	Municipality
Policy officer	M6	Municipality
Journalist	J1	Investigative Journalism Outlet
Journalist	J2	Investigative Journalism Outlet
Program manager	S1	Civil Society Organization
Head of Data Science	S2	Civil Society Organization

Table 4.1: Interviewees' encodings, along with their respective roles and affiliations

4.1 Municipality

Throughout our research on the integration of algorithms in municipal governance, we conducted interviews with key stakeholders. These interviews provided invaluable insights into the challenges, perspectives, and themes surrounding this complex topic.

One prominent theme that emerged is the recognition of the complexity involved in addressing algorithmic governance. M1, expressed the weightiness associated with the subject, emphasizing that it is a big challenge to overcome.

Transparency and accountability were also key themes that resonated across the interviews. M1 stressed the significance of ensuring transparency in how algorithms are developed and deployed. They think their organizations have not done enough to document how their systems work adequately. M3 also recognized the significance of transparency, mentioning a glass-box model as their development approach. As the name suggests, a glass-box model shows what transpires in an algorithm for those who are interested. This aims to build trust and enhance accountability.

Another central theme that emerged is the need to engage with citizens in discussions regarding algorithmic governance. M2 emphasized the importance of including laypeople in the conversation. They also highlighted the significance of the consequences of these systems on their lives. Similarly, M4 and M5 discussed the limited engagement of citizens in the decision-making process. M5 pointed out that citizens are not provided information keeping them out of the loop. They nuanced this by indicating that this is partly due to the lack of interest from the public.

Building trust between the municipality and its residents was also a significant factor mentioned in the discussions. M3 addressed the difficulties they encounter when trying to establish trust with the residents when it comes to discussing algorithms. Negative framing, fueled by framed journalistic pieces, creates unwarranted concerns and hinders progress.

Additionally, the role of digital inclusion in algorithmic governance was a notable theme. M1 highlighted the ambition to create a civic AI lab in their municipality, aiming to bridge the digital divide. By actively working towards digital inclusion, the respective municipality acknowledges the importance of ensuring equitable access to the benefits and participation in algorithmic decision-making processes.

4.2 Journalists

In our research on algorithmic governance, we had the opportunity to interview journalists who worked on the pieces [81, 82, 32] that cover the welfare algorithm in Rotterdam. They also provided us with valuable insights they gained through their investigative work and overall expertise on the subject. We approached them to help us in the search for a viable methodology for this study.

J1 suggested asking simple and straightforward questions on how they perceive things, such as how people feel about being profiled. By asking these questions, ethical and legal boundaries can be contextualized for developers.

J1 highlighted the involvement of three key parties in algorithm development: technical experts, legal practitioners, and ethicists. These parties create the algorithms, ensure compliance with legal standards such as the Dutch GDPR (known as AVG), and provide ethical guidance respectively. However, they believe citizens should be added as a fourth party. They recognize the valuable perspectives and insights they can provide. J1 also noted that fixes can be made for technical issues in an algorithm, but the structural biases that instigate these technical issues are far more challenging to solve.

In terms of awareness and involvement, J1 emphasized the need to create awareness at all levels of society. It is crucial to engage not only technical experts and policymakers but also the wider public, including elitists who may perceive themselves as unaffected by these issues. Recognizing that marginalized communities are already involved due to the direct impact they experience, inclusivity and representation of diverse perspectives become imperative to ensure equitable and just algorithmic governance.

J1 also discussed the idea of using the same data to provide services instead of referring to the 'Be Smart Start' [106] initiative. By leveraging existing data sources, they deliver efficient and effective services while respecting privacy.

Shifting the focus to J2, they highlighted the concerns and perceptions of affected individuals regarding algorithmic systems. They mentioned that affected individuals often feel depersonalized, seeing themselves as mere numbers in the eyes of machines. This depersonalization contributes to a social stigma that these individuals experience, potentially leading to further marginalization and a sense of powerlessness.

J2 also emphasized that people feel like their contribution weighs too little to matter which discourages them from active participation in discussions and decision-making. To address this issue, it is essential to empower individuals to believe their voices matter and can be of great influence.

4.3 Civil Society organizations

We have also had the opportunity to interview individuals working in civil society organizations such as initiatives and foundations focused on digitalization, digital inclusion, and AI governance.

In emphasizing the significance of algorithmic governance, S1 shed light on several crucial aspects. Firstly, they underscored the importance of cultivating awareness among individuals involved in decision-making processes. Currently, decisions are often made by individuals who lack firsthand experience of the actual impact at hand. S1 called for the inclusion of affected individuals, recognizing their invaluable insights in shaping policies and practices.

Additionally, S1 advocated for safeguarding and empowering laypeople, enabling their meaningful contributions without the requirement of extensive technical expertise. By creating an environment that supports and values the perspectives of non-experts, the potential for diverse and inclusive participation can be realized.

S2 offered a distinct perspective on the use of algorithms within their foundation. They explained that their organization employs algorithms and data obtained from the Dutch Central Bureau of Statistics to identify potential threats. However, their goal differs from for example traditional fraud detection approaches that try to catch criminals. The foundation aims to proactively assist individuals by providing them with the necessary information, thereby for example preventing mistakes and guiding teenagers away from paths that may lead to criminal behavior. The focus is on proactive intervention rather than punitive measures after the fact.

5

Methodology

In this section, we describe how we formed our study design to attain the research objective and answer each of the underlying research questions. This study design has been approved by the Human Research Ethics Committee (HREC) of the Delft University of Technology (Appendix C).

The goal of this research is to explore a different way to gather input from a broader range of stakeholders, namely the citizens. With our method, we aim to overcome some of the challenges mentioned in section 3.5.

Due to constraints in resources such as time, we altered our approach so that we can get the most out of our conversations with these often excluded stakeholders. We also made the conscious choice, on the advice of interviewee M5 (Table 4.1), not to involve the municipality yet. To this end, we conducted this study independently, without any other organization, to maintain an informal, safe, and comfortable space for these participants to voice their opinions.

Since one of our research questions pertains to community engagement, it is crucial to facilitate at least two distinct moments of engagement with the participants, with a time interval of at least one day. The first moment would serve as an introduction where presenting information to the participants is the core part. The second moment would be a discussion where we try to extract the insights and perspectives we are looking for.

5.1 Participant selection

The goal of the participant selection is to include citizens in the conversation that are generally underrepresented in government and civic engagement. By seeking out these citizens, we could gain valuable insights and perspectives that would otherwise be overlooked. These people would incidentally also be harder to reach for the public sector, considering there is some sort of distrust [53].

In light of the perceived shortcomings of the current citizen panels in achieving the desired level of success, we have made a deliberate choice to diverge from the conventional approach in our participant selection. In line with the critique expressed by interviewee M1 (Table 4.1) regarding the dominance of white participants in citizen panels in their respective municipality, we have decided to shift our focus toward marginalized and underprivileged citizens in our own selection. This way we could potentially counterbalance current biases and, in the future, gain a clearer understanding of the extent to which they differ.

To this end, we decided to go to one of such neighborhoods. Since we want to secure the privacy of the participants, we will not specify which one. What we can tell is that this neighborhood consists of mainly ethnic minorities [107], where the majority is of Turkish, Moroccan, or Surinamese background.

[44] recommends grouping citizens by interests so that they can more easily interact with each other. Instead of creating a new panel, we decided, in the interest of time, to visit local community centers and talk with the visitors.

This is possible since there are certain reserved moments when citizens come to socialize. These moments would for example contain workshops, playing cards, or having a conversation with us. This creates a comfortable and safe space for the participants to share their insights and perspectives.

We hoped to get a better outreach than the government since we were already familiar with the neighborhood. We also intentionally opted for in-person discussions to address the logistics of our objectives, considering that the conventional method of flyering, as criticized by interviewee M2 (Table 4.1), is ineffective.

It is also worth noting that participants are not always willing to share power and to open up to an evaluator [108]. It is therefore important to gain their trust [62]. Having an ethnic background and coming from the same neighborhood would make it easier for us to establish the necessary trust and understanding with the participants.

5.2 Preliminaries

It is to be noted that before actively engaging with the participants, they would have to sign an informed consent form (Appendix A), which has been reviewed and accepted by the HREC of TU Delft (Appendix C).

5.3 Part 1: Introduction

The initial session with the participants would serve as an introduction for both us and the participants. It would provide us with an opportunity to get to know the participants, establish a connection, and gauge their familiarity with algorithms. We would introduce them to our research and experiment by means of a presentation.

The presentation is divided into three parts: an introduction to the concept of algorithms, an exploration of the challenges in algorithm development, and an engaging interactive part.

5.3.1. Concept of algorithms

During the first part, we provided a clear definition of an algorithm and its origin. To enhance participants' understanding, we utilized various examples. These examples involved cooking recipes as a tangible form of an algorithm, a route planner as a straightforward illustration of an algorithm, and online advertisements as a less apparent manifestation of algorithmic processes.

5.3.2. Challenges

The second part incorporated the challenges or aspects related to algorithms. We specifically addressed data, ethics, and transparency. In the case of data and ethics, we utilized the same three examples (cooking recipe, route planner, and advertisements) mentioned in the first part to maintain consistency with the algorithms they were already familiar with. We elaborated on the role of data in each algorithm and the associated challenges. Additionally, we discussed the ethical considerations that need to be taken into account, providing relevant examples for each algorithm. Regarding transparency, we highlighted the complexities of determining the appropriate level of disclosure. As an extreme form of transparency, we presented a random piece of code to demonstrate the potential confusion and the risks of misinterpretation or biased framing that can occur.

5.3.3. Interactive segments

The final interactive part serves to ease the participants into voicing their thoughts and to prepare them for our next session, where we will be dealing with dilemmas and sharing perspectives. We have three interactive segments.

Trolley problem

The first segment is the well-known trolley problem. The trolley problem is a moral scenario where a person must decide between two choices involving a runaway trolley. One option is to do nothing and allow the trolley to continue its path, potentially harming multiple people. The other option is to actively intervene by diverting the trolley onto a different track, but this action would result in harm to a single person. It raises questions about the ethics of causing harm to save others and explores the complexities of moral decision-making. With this, we are able to refer back to the ethics part, and we will see how the participants would react if there is no 'right' answer, and what their reasoning is.

Collaborative filtering

The second segment is a prediction algorithm where different people have given ratings to various types of food using a star system, as shown in Figure 5.1.

			
Bob	☆	☆☆	☆
Jan	☆☆☆☆	☆☆☆☆	☆☆☆☆☆
Elise	☆☆☆☆	☆☆☆☆	**???**

Figure 5.1: People's rating of food

In this example, participants are asked to predict the rating Elise would give for a piece of cake. They are likely to assume that Elise would give the same rating as Jan. However, we

reveal that the resulting rating is 2 stars, which contradicts their logical expectation. This demonstrates how errors can occur in predictive algorithms, like those used in predictive policing, and highlights the challenges in avoiding such mistakes.

Classification

The third segment involves a classification algorithm that distinguishes between two distinct animal species: dogs and cats. This algorithm uses the height and tail length of the animals as determining factors, as depicted in Figure 5.2.

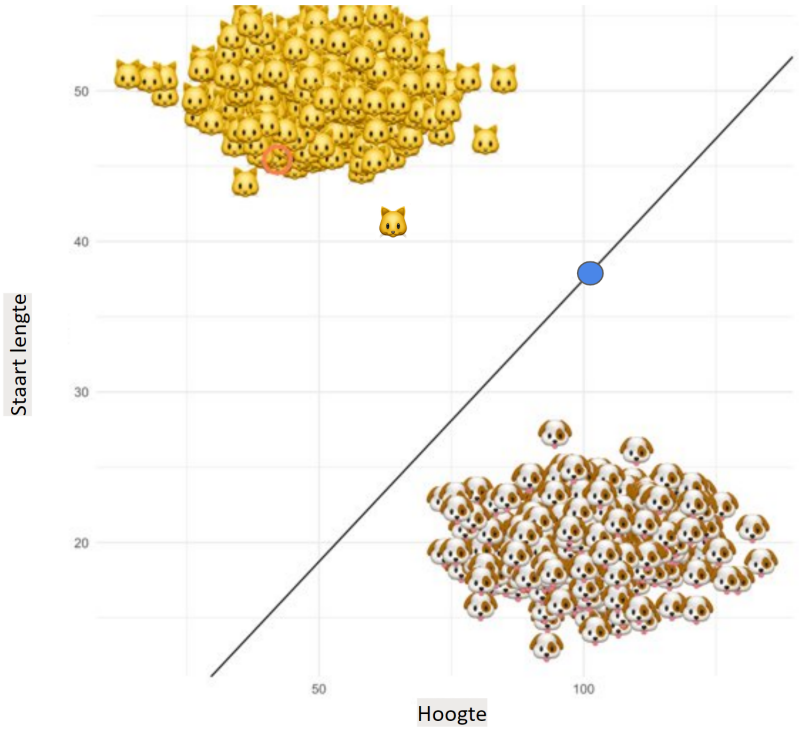


Figure 5.2: A classification algorithm with a dividing line between dogs and cats.

The dividing line separates the two species, and the blue dot represents an animal of interest. Participants are tasked with determining which species the animal, represented by the blue dot, belongs to. This exercise highlights the presence of edge cases where the answer may not be obvious and demonstrates how algorithms can sometimes make mistakes.

Concluding the presentation, we emphasize to the participants the significance of algorithms in their daily lives and the need for open discussions about them, particularly for citizens who are affected by algorithms without their input. We also extend an invitation to another session where we can delve deeper into their opinions and gather their insights on the subject matter.

5.4 Part 2: Discussion

The second session aims to facilitate discussions with the participants, allowing them to share their perspectives on specific aspects of algorithms. These conversations will be recorded to ensure accurate capture of their opinions and reasoning for analysis in our results. It is important to note that non-verbal communication may not be fully captured through recordings. However, we will attentively observe and respond to non-verbal cues during the conversations to create a comfortable environment for participants to express themselves. Our goal is to strike a balance in the narratives shared and transform any potentially obstructive emotions into constructive dialogue, in accordance with [62] on participation.

5.4.1. Goal of the session

Considering the engaging discussion that took place with the trolley problem mentioned in subsection 5.3.3, we were motivated to continue this approach of presenting participants with dilemmas and observing their perspectives.

While we appreciate the suggestion made by interviewee J1 (Table 4.1) to ask direct questions, such as questions about people's feelings regarding risk scores based on personal characteristics, it does not align with our intended objective of examining how participants navigate dilemmas similar to those faced by developers and policymakers.

Our aim is to gain a deeper understanding of their reasoning process, which could provide valuable insights and perspectives for developers and policymakers. While we recognize the value of asking these 'simpler' questions and encourage their consideration in future research, we have thus decided not to include them in our current research.

While exploring various ideas, we conducted an interview with M6 (Table 4.1), who recommended using statements to present dilemmas for participants to discuss. M6 shared their positive experience with using statements to obtain meaningful perspectives from citizens. We found this suggestion appealing as it offers a diverse range of possibilities and has the potential to facilitate meaningful discussions. Another reason is due to the conversation we have with the participants being the focus of this research. Rather than having quantifiable data, we sought to find out how these people think and reason.

During our interview with J2 (Table 4.1), we were also advised to incorporate an interactive exercise that could generate a different kind of discussion. We recognized the value of this suggestion and made the decision to implement it in our approach.

5.4.2. Statements and an interactive exercise

To this end, we prepared statements, and an additional interactive exercise called the 'Algorithmic Dashboard', of which we will provide details in subsection 5.4.5.

The statements consist of 7 pairs (14 total). Each pair consists of a question about food and a question related to a governing body, of which the reasoning we will provide in subsection 5.4.3. The participants would have to answer whether they agree or disagree with the said statement followed by the explanation of their choice. They are also provided with the statements on paper so that it is easier for them to keep track of the system and as a small reminder of their answer. The provided paper can be found in appendix B.

The idea is to use the answers to the questions as a basis to try to answer more complex questions that we are unable to ask in such a short time frame. To do this, each pair of statements tries to give a specific insight that could explain how the participants would react when dealing with more complex matters like developers have to deal with.

The discussion is kept open, so the participants are free to answer or to keep to themselves. We are aware of the warning [20] provided that inclusion is not necessarily participation, but we wanted to make the participants feel as comfortable as possible. They are also free to discuss with one another as long as it remains in the context of the subject. Otherwise, we would try to redirect their attention to the statements. Based on the answers we might also ask follow-up questions that would further specify their perspective in case we deem it necessary.

5.4.3. Food and government

The reason we went with food as a topic for the statements is because it is a topic that most people can openly talk about. Everyone has a connection with food, and it lightens the conversation for subjects that people are less willing to talk about. The other statement of each pair is mainly about the government, specifically, the Dutch Tax Administration. This is understandably a heavier subject and could put the participants on edge and less willing to cooperate. So by alternating between food and government, we aim to maintain a relaxed environment where people are willingly sharing their valuable perspectives.

The reason we mainly opted for the Dutch Tax Administration is that we expect most participants to be familiar with them and to be familiar with the childcare benefit scandal [31] in which the Dutch Tax Administration plays a central role.

The questions related to food can also serve to start off with a simplified version of the question we would like to ask about the government. This question can thus provide insight and a base understanding of what direction we are going with the following question about the government.

Food and government offer a duality of perspectives. Exploring questions about food allows participants to maintain a sense of control, mirroring the way developers create things. On the other hand, questions about the government focus on how participants perceive others' actions, allowing them to assume the role of citizens again.

While these topics are not closely related, we still think that combining them will help tremendously in drawing out their perspectives and deriving their choices for dilemmas in real-world applications that developers and policymakers face.

5.4.4. The prepared statements

We will now talk about the statements that were used. Each pair will be written followed by an explanation of both their correlation with each other and the insights we hope to gain by discussing them. The actual statements are in Dutch and were provided to the participants in Dutch but will be translated to English in this section for readability. The original Dutch version can be found in Appendix B. We would also like to note that these connections and the insights they might provide are not based on any fact but are also part of the experiment.

Pair 1

The first pair of statements consist of

- "I prefer to make my food tastier, rather than healthier."
- "The Tax Administration may use even more private data about us to better detect fraud."

This pair of statements aims to find out how far people are willing to go to get what they want. Both show two outcomes that we aim to maximize, but are presented in such a way that increasing one comes at the price of decreasing the other.

In terms of food, we aim to maximize its tasty flavor, but we also aim to maximize its nutritional value. By asking if the participants prefer flavor over health, we try to find out how much they actually value the subjective benefit over the objective benefit. The same is, in a way, done with the question about the Tax Administration.

As a law-abiding citizen, it is desirable for the Tax Administration to enhance fraud detection while safeguarding the confidentiality of personal information. However, the subjective and objective benefits can vary depending on individual perspectives. From the viewpoint of the Tax Administration, the subjective benefit lies in fraud detection, while maintaining citizen privacy represents the objective benefit. Conversely, citizens may perceive their privacy as the subjective benefit, with optimal fraud detection being the objective benefit. In our discussions with citizens, we will focus primarily on the latter perspective while acknowledging the importance of the former viewpoint.

These statements can thus provide insights into how people would decide on the trade-off between the subjective benefit and the objective benefit, maximizing one whilst sacrificing the other.

Pair 2

The second pair of statements consist of

- "I find it more important that my food looks good than it tastes optimal."
- "I think it is good enough if the Tax Administration mainly shows positive results."

This pair of statements seek to find out how valuable the positive presentation of the result is compared to the actual result. The concept that ties to this is the positive framing bias. When people present their own work to others, they often emphasize the positive aspects and play down or even ignore the negative aspects, while they are equally important.

When looking at food, the first thing we notice is how it is presented to us. We would generally rather eat something that is well-presented and tasty looking than something that looks rather unpleasant. People might as a result put too much of an emphasis on how it looks rather than how it tastes. In this case, it could even be argued that the taste is more important than the presentation of the food. The Tax Administration has a somewhat similar trade-off.

The public's trust is also important for the Tax Administration. Presenting positive results, for example, in their fraud detection algorithm helps to maintain this. However, negative results are also of high importance, since they show that there is room for improvement. Showing the negative results is also a must because only showing positive results will make

the citizens suspicious and it could all blow up in their faces, once the negative results come out as well.

We, therefore, examine the perspectives of the participants regarding the positive framing bias.

Pair 3

The third pair of statements consist of

- "I find it important to use precise quantities while cooking/baking."
- "I think it is important that the Tax Administration uses precise data about people (and not an estimate)."

This pair of statements attempt to find out the importance of precision and exactness. In both statements, it is important to realize that the cost of ideally using exact measurements and values is the investment you have to do to maintain this notion.

When it comes to the food statement, opting not to invest in measuring tools like a scale or measuring cup may lead to a more tedious process of manually measuring each ingredient. However, there are two considerations when choosing not to use precise quantities. First, it is important to evaluate how close the approximate measurement is to the precise measurement. For example, if we are slightly off by a few grams of flour out of the required 500 grams, it may not have a significant impact on the final outcome. The same can be said for the impact of a slight mismeasurement of olive oil in a tomato sauce for pasta. Secondly, the importance of each ingredient also plays a role. These two factors collectively determine whether the outcome is close to optimal or potentially a failure. A similar dynamic applies to the usage of precise data by the Tax Administration.

Using precise data is integral to algorithms, but missing data points can also occur. In this case, an estimate can be used by combining historical and similar data points. It is, however, important to note that we are not talking just about any algorithm, but an algorithm used by the Tax Administration. Inaccuracies in this context can have detrimental consequences. Therefore we would like to know what people think of using estimates in such delicate situations.

These statements can thus provide insights into how important people think the notion of precision is in different situations. It also aims to find out what level of precision would be adequate enough if optimal precision is not feasible.

Pair 4

The fourth pair of statements consist of

- "If a recipe already works well, there is no need to try something new with it."
- "If an algorithm for detecting fraud is already working well, it does not need to be reviewed/updated anymore."

This pair of statements aims to find out how willing people are to change to pursue innovation and adaptability, possibly sacrificing stability and efficiency. Both of the statements act on the premise that the current method is already good but seek to find out a better result.

With food, it can be argued that trying something new can result in a new method that makes the food tastier. We follow up on this notion and try to find out if people are willing to try that at the risk of failure. There are of course also different gradations in the changes made to the already existing recipe.

Fraud detection by using an algorithm has the same kind of notion but can have extra reasoning as to why trying something new could result in better results in the long term. Just like in food, changing the algorithm might make it perform better. It might also work in the opposite direction, but we assume that they have systems in place to detect this before deploying such an algorithm.

Unlike food, there is an adversary. The adversary would be fraudulent people. They might alter their ways and strategize based on the existing rigid system or algorithm. If the algorithm is outdated and has been reverse-engineered by the fraudsters, then they can adjust their strategy to avoid getting caught.

These statements can thus provide insights into how people value consistency and efficiency, compared to innovation and adaptability.

Pair 5

The fifth pair of statements consist of

- "I prefer to use slightly more expensive products to make my food taste a little better."
- "I think it's better if the Tax Administration uses a method that takes longer and therefore costs more money to achieve slightly better results in detecting fraud."

This pair of statements seek to find out how much people are willing to invest to have a better outcome. Both statements are under the assumption that spending more resources would lead to better results. This is somewhat similar to the trade-off that is discussed in the first pair but with a pre-emptive conscious sacrifice to obtain the desired result.

Using more expensive products can be interpreted in two ways. Using a more expensive version of a product and using a more expensive product as a replacement. Both are a monetary investment in the taste and can have a better nutritional value as a bonus.

There are many different ways to enhance the detection of fraud, but each of them will require some sort of investment. This investment is also indirectly done by the citizens. They can be seen as very small shareholders. The question to the people is thus whether they agree with their investment to be used for this.

These statements can thus provide insights into the cost-benefit analysis that the participants conduct. Showing how people would decide whether they are willing to even invest in something that can have beneficial results.

Pair 6

The sixth pair of statements consist of

- "I don't mind that there are companies like Coca-Cola that don't disclose their recipe."
- "I don't mind that there isn't full transparency about algorithms/technology used by the Tax Administration."

This pair of statements aims to explore people's perspectives on transparency and the perception of transparency, as well as their understanding of confidentiality, security, and their thoughts on information asymmetry.

In the case of food, there are companies like Coca-Cola that keep their recipe confidential. The people would probably like to know what they actually consume, but we will find out if they understand why these companies keep things close to their chest and if they agree with their choices.

The Tax Administration also needs to keep a lot of information on their technologies and algorithms secret, because of the notion we mentioned earlier that this could be abused by people that want to circumvent these systems and cause harm. We understand that it is a given that people would like to know about the systems in place up to a certain point. That is why we want the people to think about the Tax Administration being fully transparent instead.

These statements can thus provide valuable insights into people's sentiments regarding information asymmetry related to matters that impact them. They will also shed light on the level of empathy and understanding the participants have towards the parties.

Pair 7

The seventh pair of statements consist of

- "I don't mind if food is made by machines (ready-made meals) instead of by a person."
- "I prefer fraud detection to be done by humans rather than machines (algorithms)."

This pair of statements aims to find out what people think of the rapid growth of technology and the introduction of machines replacing humans. It aims to find out the acceptance of automation and trust in machines. Both statements are about an aspect where humans are slowly being replaced by machines and try to find out about their experiences and sentiment.

Ready-made meals are becoming increasingly popular nowadays. These meals are mainly made in factories and machines have a significant role in these factories. We want to see how people feel about this and if they have any experiences with these kinds of meals.

Administrative work such as fraud detection has also been increasingly automated to improve efficiency. We want to know how people feel about automation and whether they trust the decision of a machine over the decision of humans. Humans are also biased and flawed resulting in mistakes, but the real question is who they would trust more.

These statements can thus provide insights into the participants' acceptance of automation and trust in machines or algorithms.

Disconnect

We do have to reiterate that the statements that we have covered are an oversimplification of the actual dilemmas and trade-offs that developers and policymakers have to deal with.

While the subject of food and cooking can provide great insight into the perspectives of participants it is important to note that the stakes are generally much lower. Framing our statements in the context of the government helps to get closer to the applications

we eventually want to discuss. However, these still do not come close to the impactful algorithms that decide who receives basic welfare and who does not.

Therefore, while these examples help frame a discussion, it is important to acknowledge the limitations of these contexts when discussing actual algorithms that are used by the public sector.

5.4.5. Algorithmic dashboard

Following the discussion of the statements, we conclude with an interactive final discussion accompanied by visual aids. Participants are presented with a web page, depicted in Figure 5.3, which serves as a dashboard for a fraud-detecting algorithm applied to a hypothetical population. The algorithm's behavior can be adjusted using a slider, with higher positions indicating a stricter adherence to an arbitrary formula. On the left of the slider, two bars represent two groups: individuals correctly identified as fraudsters by the algorithm and innocent individuals incorrectly labeled as fraudsters by the algorithm, in other words, true positives, and false positives. The size of these bars varies based on the algorithm's strictness.

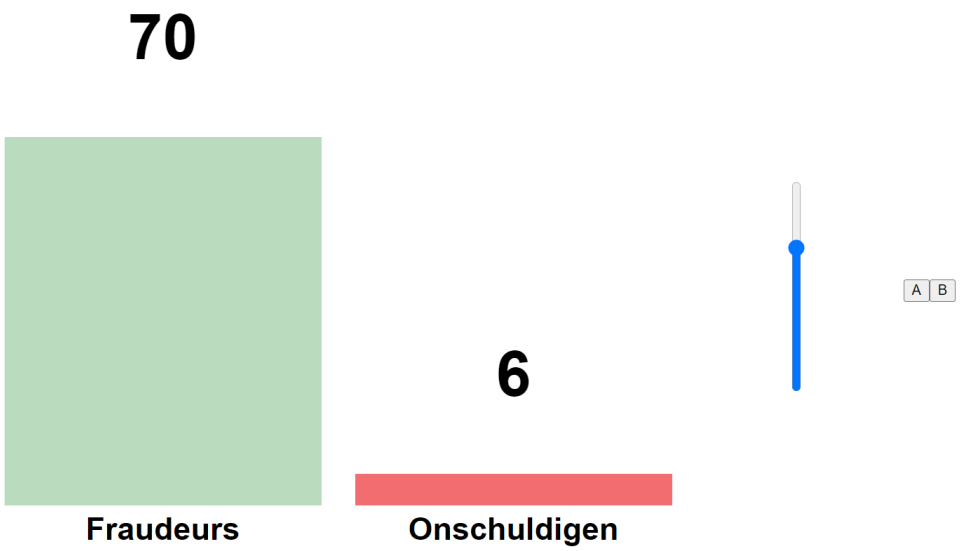


Figure 5.3: Algorithmic Dashboard in a random position. The left represents the number of fraudulent people flagged (70), and the right, the number of innocent people (6).

To begin, participants are introduced to two scenarios, A and B. In scenario A, the algorithm's strictness is set to 33%. This configuration accurately identifies 33 fraudulent individuals while avoiding any false identifications of innocent people, as depicted in Figure 5.4. Notably, this represents the highest level of strictness achievable by the algorithm without any unintended consequences, ensuring that innocent individuals are not mistakenly labeled as fraudulent.

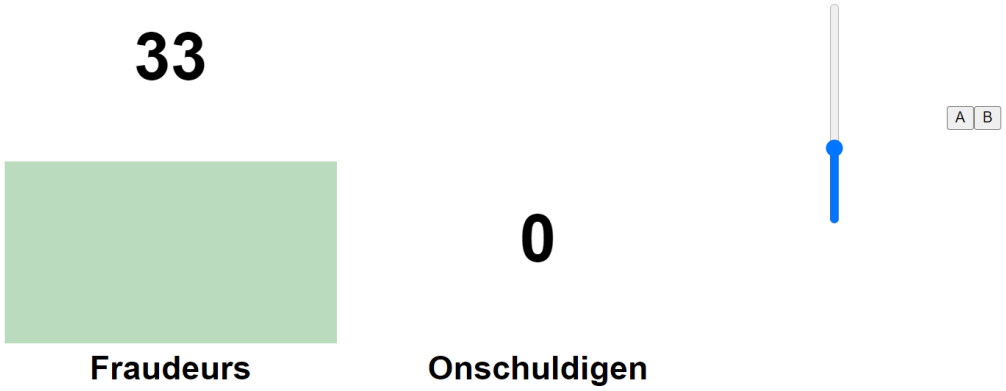


Figure 5.4: Algorithmic Dashboard in scenario A, where 33 fraudulent and 0 innocent people are flagged.

In scenario B, the strictness of the algorithm is raised to 100%, leading to the correct identification of all 100 fraudulent individuals within the population, as illustrated in Figure 5.5. However, this heightened strictness also comes at a cost, as 11 innocent individuals are erroneously identified as fraudulent.

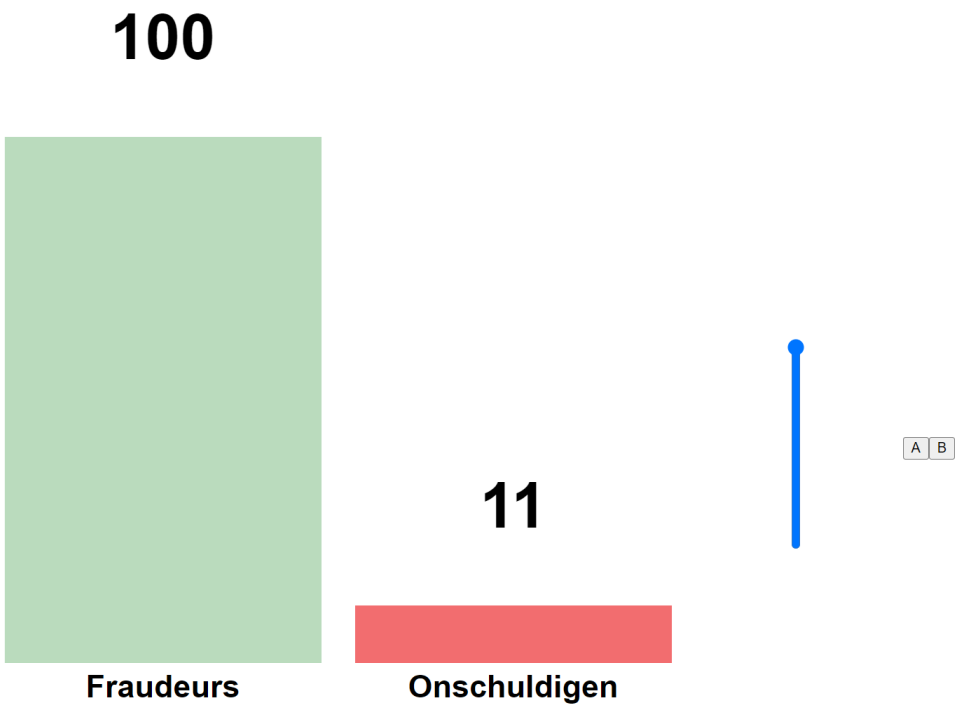


Figure 5.5: Algorithmic Dashboard in scenario B, where 100 fraudulent and 11 innocent people are flagged.

The participants have to choose which one of the two scenarios they preferred and argue why they made their choice. If there is no consensus, then we would invite the participants to discuss with each other. As a helping hand, we would adjust the outcomes by changing the level of strictness until a middle ground was reached or a consensus is met among the participants. The formula used for calculating the algorithm's outcomes would not change.

Another adjustment would be the consequences for innocent people that are mistakenly identified as fraudulent. Depending on their answer we can make the consequences lighter or more severe. This would range from getting only a letter from the Tax Administration that asks the person to report to their headquarters which is corrected within a week, to actually having to prove their innocence in a courtroom.

This interactive exercise serves as a method to explore how participants make the trade-off between false positives and false negatives in the context of algorithmic decision-making. By presenting the two scenarios, the participants are prompted to consider the consequences of these notions. The exercise allows participants to engage in a thought-provoking discussion about the potential harms and benefits associated with each scenario.

The participants can reflect on their own values, priorities, and risk tolerance, leading to a deeper understanding of the complex decision-making process involved in striking the right balance between false positives and false negatives in algorithmic systems.

5.5 Processing

Given that the discussions were recorded in audio format, we have written up transcripts of the dialogues. This process went through three rounds.

First, we transcribe the entirety of the conversations while assigning unique anonymous identifiers to each participant to ensure their anonymity. These identifiers are in the form of a letter followed by a number (e.g., ”P1 did not agree”).

In the second round, we refine the initial transcripts by confirming the validity of what was written and added additional information about the participant such as their tone, emotional state, and their specific emphasis, if applicable (e.g., ”P2 made a jesting remark”).

In the final round, we translate everything into English. During this round, we paraphrase as needed to maintain a natural flow of conversation and to make sense of sentences that do not directly translate (e.g., the saying in Dutch ”Raw beans make food sweet” translates to ”Hunger is the best sauce” in English).

6

Practical Challenges

With the experiments we have conducted explained in our methodology, we also gained some valuable experiences in the field of social interviewing and discussing. This section talks about the experiences we gained leading up to and including the discussions we had with the citizens.

6.1 Hard to find willing community centers

As mentioned in our methodology we specifically went to communities in a marginalized neighborhood as they should be considered first as stakeholders [62]. We intended to engage with these community centers directly as they are typically accessible and serve as focal points for interactions.

However, our attempts shed light on the difficulties and challenges that need to be faced to only reach out to these communities. We approached a total of seven community centers in the neighborhood. Of these seven, four expressed their willingness to participate and help in our research. The remaining centers either did not respond at all or declined to be involved referring us to the municipality instead, which we wanted to avoid as mentioned before.

Unfortunately, two of the initially willing centers eventually withdrew as well. One center faced resistance from its visitors who did not want to participate. The other center withdrew from this research due to inevitable language barriers which we will discuss in the next section.

Consequently, we were left with two community centers that fulfilled the criteria of having willing visitors, the available time, the openness to engage, and visitors who could sufficiently speak Dutch.

6.2 Language barrier

Our first successfully scheduled session took place at a community center, where we engaged with a group of approximately 20 male visitors. We were told by the administrators that this was the only group they had available for our research.

Typically, the group would gather once a month in the evening where they would cook and have dinner together. During the planning discussion, we agreed that it was best for

us to give our presentation, which is outlined in the methodology, while the participants were enjoying their cooked meals. Recognizing that the participants were non-native Dutch speakers, we consulted the administrator whether they were sufficiently proficient in the Dutch language to understand the presentation and thereafter have another session to have our discussion. We were assured that they were.

However, the evening did not proceed as smoothly as anticipated. When starting the presentation, the participants made it clear they would prefer the presentation to take place after their meal instead. In the interest of accommodating their wishes we complied. We were also invited to have dinner with them which we accepted. In hindsight, this proved to be a pleasant experience, in which we were able to get to know the participants and build comfort.

Following the meal, we began our presentation. While giving the presentation, it became apparent that the participants had difficulty trying to understand what were saying. At first, we thought the technical terms might be too much, but after adapting our language use to be more accessible to non-native speakers, we noticed that this did not change their expressions. The participants then informed us that most of them were not proficient in Dutch after all. They requested that we continue the presentation in the Berber/Amazigh language, which we regretfully explained was not possible for the current session, considering we were not proficient enough ourselves in that language. We did explain that we would consider it in the future, using a language interpreter. Nonetheless, we still proceeded with the presentation for the participants that were able to follow along.

After concluding our presentation, one of the participants took the initiative and tried to summarize our presentation in their native language so that others were also at least informed of what the presentation is about and why we were there in the first place. This demonstrated their interest in the topic and willingness to assist us, which was highly appreciated. In the interest of time, we decided against moving forward with this group.

Despite the session not going as initially planned, we still gained valuable experience, identifying areas for improvement and acquiring useful insights for future engagements.

6.3 Singular session

Another challenge was the availability of one of the groups we conducted the study with since they were only available for one session. To this end, we adjusted our methodology to maximize the value of this single session.

We decided to keep the presentation at the start of the session but reduced it significantly, ensuring that the participants were still well-informed on the subject of our research. The condensed presentation contained only essential information about what an algorithm is in the context we would discuss and highlighted some of the key aspects. These key aspects are data, ethics, and transparency. We intentionally avoided detailed examples and omitted other interactive parts, such as the trolley problem and the food rating predictions. Since we would also go over the statements we did not think these interactive parts were necessary.

Subsequently, we did proceed with all 14 prepared statements and the interactive algorithmic dashboard as explained in the methodology.

6.4 Adjust to participants

A notable observation about both groups in which we conducted the sessions successfully is that all participants belonged to the elderly demographic. This was both beneficial and challenging for us. The advantage presented was that these elderly people belong to a demographic that is even more of a minority, especially in the context of technology and algorithms. The challenge, however, was tailoring our discussion toward the elderly. This entailed several factors.

First and foremost, as expected, the elderly were not familiar with technology, rendering some of our examples less relatable to them. This made it harder to explain certain topics but reaffirmed our decision to choose food as a subject as a means for them to communicate their perspectives.

Secondly, the administrators warned us about the participants' attention span. They have difficulty maintaining focus on a subject for an extended period of time. As a result, we changed our presentation and discussion to tailor to them, spanning between 30 to 45 minutes. Unfortunately, this meant we also did not have the opportunity to discuss one topic extensively.

Thirdly, we observed some of the participants not always being present mentally. An example is a participant that even dozed off briefly during one of our discussions. There were also instances where they expressed their confusion by answering their own interpretation of the statements, which we had to decipher during our discussion.

6.5 Key takeaways

Our experiments also yielded several takeaways that will help in shaping our future approaches;

Firstly, we learned about the importance of gathering specific information prior to meeting the participants. This includes understanding the demographic profile of the participants and ensuring that the necessary tools and resources, such as a beamer, are available at the location.

Secondly, it is important to acknowledge that participant dynamics can change and that not everyone is consistently present for each session.

Finally, conducting these studies has allowed us to expand our network of people that are interested and invested in the subject. This broader reach has opened up new ways for collaboration with other parties.

6.6 Prejudices

We also think it is noteworthy to mention the stance of the participants towards the government and the Tax Administration in particular during our discussions. They felt like the Tax Administration violates their privacy all the time and is only out to harm by taking people's hard-earned money, mentioning "The big brother watches you" and "Everyone is afraid of the Tax Administration".

Another individual expressed their discomfort with the topic, saying “I get chills from taxes, I do not like talking about taxes”. This amplifies the overall negative sentiment toward the Tax Administration. One of the participants got emotional when discussing unlawful accusations, stating “To be accused of something you did, it’s not pleasant!”.

Overall these quotes provide a glimpse of the experiences and attitudes of the participants toward the government and the Tax Administration in their handling of taxes and fraud accusations.

6.7 Their experience

On a more positive note, we have received several enthusiastic responses as well on the research we are conducting. They found the sessions to be educational and informative. They thanked us a multitude of times for presenting the information the way we did, saying “You explained it very well and kept it fun!”, and even asked for us to return.

7

Results

In this section, we will present the resulting discussions that took place based on subsection 5.4 explained in the methodology. We would like to reiterate that these were open discussions. We wanted to make the participants feel as comfortable as possible. To this end, each participant was free to voice their opinion or to refrain from participating if they choose to do so.

7.1 Groups

We have been able to do the experiment with a total of two groups, which we will call Alpha and Beta respectively.

7.1.1. Alpha

The Alpha group is the group mentioned in subsection 6.3 with whom we were only able to do a singular session. The group comprised a total of 7 participants, with 6 actively engaging in at least one discussion during the sessions. All participants in the group were elderly. Among the active participants, one was of Dutch descent, while the remaining individuals were of Surinamese descent. The active part of the group consisted of 2 male and 4 female participants.

7.1.2. Beta

In the case of the Beta group, we had the opportunity to conducted both an introductory session, as detailed in subsection 5.3, and a subsequent discussion session, which will be discussed in this report. The Beta group consisted of four elderly participants, all of whom were female and of Surinamese descent.

7.2 Statements

As mentioned in subsection 5.4.5, we have encoded the participants with identifiers to retain their privacy. The Alpha group's encodings begin with the letter A (e.g., A1) and the Beta group's encodings start with the letter B (e.g., B1).

For each of the groups, we will provide a description of the discussion held regarding each statement and the key findings that emerged from the discussion.

7.2.1. Alpha group

Statement 1: I prefer to make my food to be tastier rather than healthy.

Within the group, there existed a divergence of viewpoints regarding the relative importance of the two aspects of food. A1 held the perspective that the most important factor was the nutritional value of the food. In contrast, A3 and A7 concurred that taste took precedence over health concerns, arguing that the majority of individuals primarily seek enjoyment rather than focusing on health benefits. A7 expressed this viewpoint by asserting; ”I think most people think taste is more important, not health. Most people just want to eat nice food.”

Compared to the extremes represented by A1, A3, and A7, participants A2 and A5 presented a more nuanced stance, acknowledging the significance of both taste and health. They advocated for a middle ground, suggesting that food can and should be both delicious and healthy simultaneously.

Statement 2: The Tax Administration may use even more private data about us to better detect fraud.

The consensus of the group was resolute in their disagreement regarding the statement. A2 articulated a personal inclination towards privacy, a sentiment that resonated with A1, who shared a similar perspective. Both A7 and A2 voiced the opinion that privacy had become a relic of the past, expressing sentiments such as ”Privacy does not exist anymore” and ”I would rather have my privacy, but that does not exist anymore, so yeah” respectively. As the discussion progressed, the participants were prompted with a follow-up question of whether anyone would be willing to disclose more personal information to facilitate improvements. In response, A5 asserted a strong preference for maintaining their privacy saying “No, I would rather have my privacy. A2 added to this with a remark insinuating that an individual’s personal information was already extensively known, to which everyone agreeingly laughed.

Statement 3: I find it more important that my food looks good than it tastes optimal.

Within the group, a certain degree of disagreement unfolded in relation to the statement. A1, having a preference for the visual aspects of food, articulated this by emphasizing the importance of well-presented food when saying “I think it is important for it to be neatly presented”. A2 tried testing A1 suggesting it would not have a nice taste anymore, to which A1 admitted the possibility. A2 then took the opportunity to express their views of taste being more important than presentation, stating “Well, I think that whenever I prepare something, it needs to taste nice. I will not mess it up, of course, but the taste is more important”. A1, in response, elaborated on their perspective by replying with “If food is not served well enough, then I have already finished my meal so to speak”. This reaffirms their belief that food needs to look nice in order to taste nice.

Prompted by a follow-up question, we asked them to consider a hypothetical scenario where they would have to prepare a dish for a visiting king. Both A1 and A2 united in their response, suggesting that there would be no need to alter their preparation. They also expressed their view of the king as just human as well when saying "If he is hungry, then he would eat.

Statement 4: I think it is good enough if the Tax Administration mainly shows positive results.

The participants had difficulty understanding the statement, which prompted us to rephrase the statement into a simpler question. The question we asked was "Do you mind that negative results are not really shown by the Tax Administration regarding their activities?". Most participants nodded in agreement. A2 expressed this vocally by stating "I do mind actually, they should share that [information]", to which A1 agreeingly said, "Yes [they should], but they don't".

With that, the group had a better understanding of the statement in question, so we decided to circle back to that. When returning to the original statement, A2 expanded on their viewpoint, stating "They should show both. Because you do not know how many [innocent people] they have [unlawfully] caught and what they did [to them]". A1 also provided a valuable insight highlighting the common tendency among people to focus on the negative rather than the positive, by stating "people are often inclined to focus and the negative rather the positive, you see it often on tv as well" to which was inferred "Such as the news" and they responded with "Yes, exactly" reinforcing the notion.

Overall, the group expressed a collective agreement that the omission of negative results is something to be mindful of, with A2 emphasizing the importance of transparency and A1 drawing attention to the prevalence of negative focus in various media, including television news.

Statement 5: I find it important to use precise quantities while cooking/baking.

The group's response to this statement was predominantly in disagreement, expressing they prefer cooking based on intuition or gut feeling over following recipe instructions precisely. A7 specifically mentioned that minor changes are not imperative, stating "If it is a little more or less, then it does not bother me all that much". However, A6 stood apart from the rest of the group by not derivating from the specified instructions by expressing "I follow the recipe precisely". When asked for elaboration, A6 explained that they lack the familiarity and therefore the intuition to rely on their gut feeling while preparing something but A6 also indicated this might change with more experience, explaining "It is unknown to me, so I follow the instructions. Whenever I pick up more experience, then I would be more inclined to dit based on intuition, but [until then] I use exact quantities". This contrast highlighted the diversity, where the majority relied on intuition, while A6 preferred a more methodical approach, so as not to make any mistakes.

Statement 6: I think it is important that the Tax Administration uses precise data about people (and not an estimate).

The consensus within the group was unanimous, with all participants agreeing with the statement. A2 specifically criticized using estimates, saying "If they estimate, you will still end up paying them".

When asked if their perspective on the importance of precise data meant the same to them if it also meant the Tax Administration would be aware of more information, such as price amounts spent in a supermarket and therefore potentially breaching privacy, the conversation took an interesting turn. A2 expressed their lack of concern, stating, "It's fine by me, I have nothing to hide anyway, and I can't hide anything from them either". A7 chimed in, emphasizing the extent to which the Tax Administration is already aware of detailed knowledge, mentioning their knowledge of even your bathroom knowledge as a joking remark. A2 confirmed this sentiment by agreeing, "Yeah, it's horrible".

This statement sparked another moment where participants expressed their distaste for the authorities highlighting the growing intrusion of personal privacy by authorities such as the tax administration.

Statement 7: I prefer to use slightly more expensive products to make my food taste a little better.

The group showed their culinary expertise by engaging in a discussion about the relationship between the price of ingredients and the taste they can provide. A2 challenged the assumption that more expensive products lead to better-tasting dishes, stating, "That is not the case with food. It depends on what you make" and "That is what they say" implying skepticism towards the assumption. A3 echoed this notion, explaining their approach to cooking is based on personal preference rather than focusing on the price of products, saying, "I just make whatever I feel like making. It does not have to be expensive".

The group was then prompted to a hypothetical scenario involving some merchant advising them to pay for more expensive products to make their food taste better and if they would consider taking their advice. A3 was quick to respond with a resounding "no" indicating they completely oppose to the merchant's advice. Contrary to A3, A7 expressed they are open to the idea, stating "I would at least taste it, of course". A2 proposed an alternative approach, suggesting to explore other products stating "I would try an alternative. At the end of the day, it comes down to the same thing [for me]". A3 then raised a valid point, emphasizing the ability to even work with the ingredients, explaining, "You need to be able to work with expensive products. If you can't make it work, then there is no reason to use more expensive products".

This exchange highlighted the participants' nuanced perspectives on ingredient quality, showcasing the understanding that taste is influenced by various factors beyond the price tag, including personal preference, cooking skills, and the compatibility of ingredients with the desired dish.

Statement 8: I think it's better if the Tax Administration uses a method that takes longer and therefore costs more money to achieve slightly better results in detecting fraud.

Participant A7 did not feel comfortable with the previous question regarding the Tax Administration, so we decided to skip this statement, as we wanted to maintain a comfortable space where they could be open about what they think.

Statement 9: If a recipe already works well, there is no need to try something new with it.

For this statement, all participants, except for A6, reached a consensus. Participant A3 referred back to the discussion about using more expensive products, highlighting the opportunity to try something new and different. A2 shared their adventurous approach, stating, "I do try that sometimes. If think it it is not good enough, then I just don't continue".

As said, A6 held a contrasting viewpoint, expressing a preference for consistency in using the recipe by always following what it instructs. When questioned about the possibility of wondering how a small variation might affect the taste, A6 humorously responded, "I don't have that kind of fantasy". A7 then playfully teased A6, saying, "You always need to at least try something new", to which A6 responded, "I try enough new [other] things, but that is not the topic. If I follow the recipe exactly, I know what I can expect in terms of taste".

This exchange showed the diversity in the group. A2 and A3 expressed their openness to experimenting, while A6 preferred consistency.

Statement 10: If an algorithm for detecting fraud is already working well, it does not need to be reviewed/updated anymore.

After A7 regained comfort, they humorously agreed to participate as long as the questions were not directed toward them. With that, the participants were open to discussing statements concerning the Tax Administration again.

Discussing the statement, all participants quickly agreed on the importance of measures taken for continuous improvement in evaluation even if the algorithm already functions effectively, thereby disagreeing with the statement. They reflected their understanding of adapting to evolving circumstances to ensure effectiveness.

Statement 11: I don't mind that there are companies like Coca-Cola that don't disclose their recipe.

This statement triggered a heated argument between participants A1 and A7 regarding the secrecy surrounding the recipe of food and beverage companies like Coca-Cola. A6 shared that they do not consume Coca-Cola anymore due to the lack of knowledge of their ingredients. A3 stated they don't drink it anyway, so they could not be bothered, showing they do not care as long as it does not affect them.

A1 clearly expressed their dissatisfaction with the lack of transparency regarding the contents of the recipe, emphasizing the need for disclosure by stating, "I do mind them not

sharing the contents of their recipe". In response, we posed another question to see how A1 would feel if making the contents of the recipe public, would lead to other companies replicating the beverage and in turn undermine Coca-Cola's business model. A1 stuck to their perspective, insisting on the need for disclosure, saying, "They must disclose what is inside".

A7 then interrupted with "No, no. I do not agree with that. If you designed [or created] something, then it is yours and yours alone". A1 then argued that secrecy has negative consequences, explaining, "But people are suffering because of it, society is suffering because of it". A7 then tried to elaborate on their point by saying, "Imagine you design something, and it could be worth a lot of money, and someone steals it?", to which A1 reminded A7 that the question was about personal feelings, and they personally believed that information should be disclosed. In response, A7 stated, "I don't mind, since I can understand why they do it".

A1 then brought the discussion back to the topic of nutrition and directed the focus on the impact these beverages have on society's health, emphasizing their concern. A7 stated that A1 was right in having concerns, to which both agreed that whether you agree or not, people will still continue to consume products like this.

This passionate debate where the participants expressed their differing opinions, while also finding common ground showcased the potential of having these discussions.

Statement 12: I don't mind that there isn't full transparency about algorithms/technology used by the Tax Administration.

With growing comfort, A7 felt confident enough to issue a warning statement, expressing their agreement with the necessity of fraud detection, but at the same time add a warning concerning the pervasive presence of surveillance, saying, "They should detect fraud, I completely agree. But let me tell you something about the Tax Administration and the government, the big brother is watching you. They don't have to show how, but they will do it.".

A2 then added, saying, "They have their own set of rules, but they should be more transparent towards society", emphasizing the importance of the authorities being transparent and meeting the population halfway. A2 also referred to the democratic nature of the country by saying "We get the chance to vote for them, it is a democratic country after all, isn't it?" as a means to show their own significance.

Statement 13: I don't mind if food is made by machines (ready-made meals) instead of by a person.

The participants unanimously disagreed with the statement, expressing their preference for homemade meals. A2 emphasized their dislike towards the growing technological development saying "In the future, we will be treated by machines [like babies]. I do not like them [read-made meals] anyway. At least when I prepare food myself, I know what I'm eating". When asked about having someone else prepare their food, A2 asserted that even in such cases, they are aware of the ingredients used.

A1 acknowledged that there are individuals who cannot cook and might benefit from ready-made meals. A1 and A2 agreed that these individuals could find value in such

convenient options.

To further explore this statement, we prompted the question of which they would prefer, a ready-made meal or a meal prepared by someone who is a poor cook. A7 diverged from the question saying they wouldn't know since they have never tried a ready-made meal and don't intend to do so either. A2 shared a more open-minded perspective, suggesting that a ready-made salad might not be all that bad. A2 also mentioned their daughter's reliance on these meals due to her work. With this, they also highlighted the viewpoint of many young people today that, due to time constraints, are forced to opt for ready-made meals. They also suggested they might consider a visually appealing ready-made meal in the future. A1 and A3 agreed, indicating that convenience and presentation would be key factors in whether they would consider it as a choice in the future.

Statement 14: I prefer fraud detection to be done by humans rather than machines (algorithms).

The participants engaged in a discussion about the increasing role of technology in their lives, particularly in relation to tasks such as mailing and other computer-based activities. A2 acknowledged their role saying "Computers are everything these days. Mailing etc is all done using computers, so what can we even say [to counter this]". In response, A5 expressed a preference for human involvement, trusting people more than machines to handle such tasks. A2, however, argued, "But that will take loads of time. These days everything is done using computers, so manual stuff ceases to exist", explaining that relying solely on humans would be time-consuming, emphasizing the efficiency and speed offered by computers.

A5 maintained that trust is the prevalent factor in this issue, outweighing considerations such as efficiency. While A2 agreed that trust has a higher importance, they also indicated the practicality of using machines saying, "It is important, but time is also important. You will fall behind, life goes on. Look at elections, for example, they take days to finish the count [in the Netherlands], but if you do it using a computer, you will be done quickly".

Overall, the participants recognized the importance of computers in most aspects of life but differed in their opinion of what should be regarded as more important.

Algorithmic Dashboard

The majority of participants expressed a strong preference for minimizing the risk of innocent people being falsely accused of fraud. However, A2 held a contrasting opinion, stating, "I would rather have as many people [criminals] caught. The innocent people will land on their feet, so I would rather have as many guilty people caught", leaning towards a scenario where a larger number of guilty individuals are caught, even if it means some innocent people may be wrongly accused.

A3 disagreed with this perspective by stating, "It is important to catch criminals, but accusing innocent people for no good reason is very unpleasant. I know it's not pleasant" and "Accusing someone without justification is not right; it is unjust. I do not care how little the number of criminals caught is, as long as there are no innocent caught", emphasizing the importance of avoiding unjust accusations and highlighting the negative impact it can have on individuals. A6 echoed this sentiment, expressing a desire to prioritize the avoidance of

false accusations, even if it means fewer guilty individuals are caught, to which A5 vocally agreed.

We then asked the odd one out, A2, why they think it is important that as many criminals need to be caught at [at the cost of innocent people]. To which the participant replied, "There are a lot of places where innocent people are wrongfully accused. These innocent people can always be defended in court".

In an attempt to gauge the participants' flexibility, they were asked whether their opinions would change if the consequences of being accused were less severe. A3 maintained their original stance saying, "It does not matter, nobody innocent should be accused".

Overall, the vocal majority recognized the significance of protecting innocent people, while A2's perspective focused on the positive side of cranking up the algorithms' strictness.

7.2.2. Beta group

Statement 1: I prefer to make my food to be tastier rather than healthy

For this statement, most participants expressed a preference for taste when it comes to this dilemma. B4 was initially on the fence, stating they preferred food that is both healthy and tasty. However, when we presented them with a binary choice between the two, they leaned more towards food being healthy while still emphasizing the importance of taste. B3 reiterated their preference for taste and B1 agreed. B2 attempted to rationalize their choice saying "I prefer the taste, but in my case, there are some vegetables too, so it is healthy as well".

Overall the majority of participants prioritized taste, although they did acknowledge the importance of food being healthy as well.

Statement 2: The Tax Administration may use even more private data about us to better detect fraud.

B3 and B4 both disagreed with the statement. B3 firmly expressed their stance, saying, "I think it is more important that I retain my privacy". When we questioned them about their stance if it meant that more criminals are caught as a result, B3 responded skeptically, saying, "That's what they say, but I don't know if that's the case. I mean, they use it for these criminals, but they always find a workaround. It's actually a trick since the criminals get away with it anyway".

In contrast to B3, B4 took a different perspective on their privacy stating their innocence saying, "I have nothing to hide, I pass everything along [information]. They will find out anyway". However, with this last sentence, the discussion shifted toward the participants' perception of the Tax Administration, with B4 remarking "Everyone is scared of the Tax Administration" and "I'm scared myself". B4 also expressed their surprise that, despite this seemingly common notion, some individuals still engage in fraudulent activities by saying "And there are still people that commit fraud anyway".

The general consensus of the group was that receiving a mere letter from the Tax Administration would already cause fear and concern. The participants mentioned they would rely on their children, highlighting the anxiety and lack of understanding they experience when receiving a letter.

Statement 3: I find it more important that my food looks good than it tastes optimal.

B2 and B4 expressed their belief that both taste and appearance are equally important when it comes to food, with B4 also emphasizing the importance of food being healthy and B3 agreeing. B4 provided an example, mentioning "Sometimes food looks tasty, but it actually isn't [tasty]. That is the case in restaurants for example".

To probe them further we prompted them with questioning which of the two they would prefer if it came down to it. B4 responded by stating "I would rather have it not look as good as it tastes. But yeah, if you paid then might as well eat it anyway" showing they prefer taste over presentation. In contrast, B3 expressed a different approach, stating they would not even eat a meal they do not particularly find tasty, saying "If I don't like the taste, I won't eat it. I would order something else, no matter how much longer I have to wait".

B2 was skeptical of their stance replying with "Hunger is the best sauce" to indicate B3 would still eat the distasteful meal if they were hungry enough, to which B1 wholeheartedly agreed. B3 stood firm in their opinion, continuing to emphasize their preference for hunger over eating food they do not enjoy.

Statement 4: I think it is good enough if the Tax Administration mainly shows positive results.

B3 expressed their disagreement by expressing "I think they must also report negative results as well". B4 presented a counterargument, suggesting that negative outcomes are already highlighted and emphasized enough in the media, expressing their distaste for the media supposedly bloating up some cases of negative outcomes. B2 supported this view by noting, "They [the media] only start talking whenever something is wrong".

We then reformulated the initial statement into the following question: "Do you think it is important to have an equilibrium of negative and positive results?". In response to this question, the entire group agreed, expressing the belief that there should be a balance between negative and positive reported results.

Statement 5: I find it important to use precise quantities while cooking/baking.

B3 agreed with the statement, but with the condition that it applies when they are making something for the first time stating, "If I make something for the first time I make sure to look at the precise ingredients. But afterward [not anymore]", suggesting once they become familiar with the recipe, they rely more on intuition and gut feeling. B4 and B2 agreed with this approach, adding "Then you'll know [how much to use]".

B3 elaborated, mentioning it depends on the situation. "If I'm baking cookies, then I'll always weigh the ingredients otherwise it always fails. But with cooking regular food, for example, vegetables and such, I just do it based on gut feeling", shows how they would proceed depending on the situation.

When specifically asked about their opinion, B2 mentioned that they cook at their own pace and stated that they do not use recipes, emphasizing with "I just do everything based on intuition". When we insinuated that B2 is an experienced cook, they replied "Maybe, maybe not" and acknowledged that others may not like the food that B2 likes.

Statement 6: I think it is important that the Tax Administration use precise data about people (and not an estimate).

The participants expressed their prejudices and negative experiences with the Tax Administration, believing that the authorities already know everything and make mistakes. B4 emphasized this by saying, "They often make mistakes as well" and provided an example saying "They give you some money, which makes me happy, but I end up having to pay them back [because of their mistake]". B3 agreed, highlighting that mistakes only seem to occur when it comes to paying them.

To track back to the statement, we presented a hypothetical scenario where estimates would be used and asked the group about their stance regarding this. The participants unanimously agreed on precise calculations as the preferred option. They emphasized this by saying "That would make it easier for us. [That would give us] no anxiety, no stress". So when it came down to it, the participants expressed their desire for accuracy and precision when dealing with non-trivial matters.

Statement 7: If a recipe already works well, there is no need to try something new with it.

B4 expressed that trying out different ingredients doesn't have to be the norm saying "That does not have to be the case", but quickly admitted that they do it sometimes. B3 shared that they often experiment stating "I do actually, I often try different herbs and spices". B1 acknowledged that it might make the food tastier, but remained undecided. B2, on the other hand, strongly disagreed, stating "I would not do that. If a recipe works well then it's good enough. If you perform well on your job, then why would you change that". B1 tried to reason saying that it might taste better".

The participants that were in favor of trying new methods or ingredients, were prompted with the question "Does it ever go wrong, and if so, does that change your perspective?". B3 responded by acknowledging that sometimes things can go wrong saying "Oh yeah, of course, that happens", but in such cases, they simply choose not to eat the dish and explore other herbs and spices available in the market. B4 then anticipated the next statement when explaining they why don't think changing a recipe is necessary, saying, "If the Tax Administration doesn't make any mistakes, then we would not change that either".

Statement 8: If an algorithm for detecting fraud is already working well, it does not need to be reviewed/updated anymore.

B4's answer was already known due to their stance on the previous statement. The rest of the group also agreed. B4 further expressed that revising the algorithm is not necessary.

To challenge this, we sketched a scenario where fraudulent people would adapt to the algorithm, suggesting that keeping the algorithm the same results in fraudulent people being able to bypass the system. Their perspectives changed and they acknowledged the need for changes. B4 expressed "In that case, I think you have to make a few adjustments" to which the rest of the group agreed.

Statement 9: I prefer to use slightly more expensive products to make my food taste a little better.

B1, B3, and B4 initially disagreed with the statement suggesting that more expensive products are not necessarily better. B3 emphasized that cheap products can be just as good saying "I don't use more expensive products. Cheaper products are just as good". However, B2 presented a different perspective based on personal preference and cultural influences. They explained "In my case, I use Surinamese products, and those tend to be more expensive. If it is about the food that I eat and those are more tasty. I, for example, wouldn't use spinach in my food if snake bean is available". B4 also acknowledged that certain more expensive ingredients, such as snake bean, can be worth the extra cost.

B2 further emphasized their willingness to pay a bit more stating "I mean in this case I would rather have more expensive products. I am willing to pay those extra few euros to get something that I like. I always use exotic Surinamese products and those are a little more expensive, but yeah, I like it". B1 and B3 were eventually convinced by the arguments presented, with B3 noting their own use of Moroccan and Turkish herbs and spices that may fall into the category of more expensive products.

Statement 10: I think it's better if the Tax Administration uses a method that takes longer and therefore costs more money to achieve slightly better results in detecting fraud.

All the participants in the discussion agreed with the statement that it is better to invest more time and resources if it leads to improved results. B3 expressed "If it is already used [money], then why not for this. If this means it will work better I don't see why not". B2 supported this perspective by stating "It is okay to take more time, as long as it actually gets better results". Additionally, B3 argued that if the enhanced measures lead to improved efficiency and reduced losses on the Tax Administration's side, it would ultimately benefit the citizens who would no longer have to compensate for those losses.

The participants' agreement highlights their belief in the importance of investing in systems and processes that deliver more effective and efficient outcomes, even if it requires additional resources or time.

Statement 11: I don't mind that there are companies like Coca-Cola that don't disclose their recipe.

B3 demonstrated an understanding perspective by acknowledging that Coca-Cola, as the manufacturer, has the right to keep its recipe a secret arguing "I understand why they wouldn't share [their recipe]. I mean it's their product". B4 added "Otherwise nobody would drink it" explaining that if the recipe were publicly known, it might discourage people from consuming the product. B2 agreed, stating that Coca-Cola has no choice but to keep the recipe confidential to protect its brand and maintain its business model.

When asked if they were not bothered by the fact that there might be ingredients they don't like in the beverage, B4 humorously responded "That's why I don't drink it". B3 further expanded on the idea, highlighting that Coca-Cola is not the only product with undisclosed ingredients and that this is a common practice in the industry elaborating with "There are many more products of which you don't know what you are actually consuming,

so yeah, to each their own".

This discussion highlights the participants' acceptance of the fact that companies have the right to keep their recipes and ingredients confidential.

Statement 12: I don't mind that there isn't full transparency about algorithms/technology used by the Tax Administration.

B3 and B4 expressed agreement with the statement, emphasizing the importance of secrecy and stating that Tax Administration should keep their method a secret. They believed that if all the information is revealed, it could potentially lead to negative consequences saying, "If they will tell everything, then something [bad] might happen again".

However, B2 held a different perspective, expressing a desire to know more about the way the Tax Administration operates. Participant B2's viewpoint influenced B4, who then changed their mind and agreed that having a little more information on the matter would be desirable.

This exchange reflects a diversity of opinions among the participants, with some valuing the need for secrecy and others expressing a curiosity for greater transparency.

Statement 13: I don't mind if food is made by machines (ready-made meals) instead of by a person.

All participants expressed their preference for freshly made meals over ready-made options. They emphasized the importance of taste, quality, and the satisfaction of preparing food themselves. B4 specifically mentioned, "If I make roti [Surinamese dish] then it is always tasty, it might take longer but it is always tasty", to which B2 agreed. B3 added explaining they never buy ready-made meals. B4 also expressed their dislike for frozen food, associating that with ready-made meals, and their preference for freshly made food.

B1 also emphasized the preference for food made by humans rather than machines. However, B2 shared an experience of eating a ready-made meal at a friend's place and acknowledged that they would consider it if necessary. B4 agreed with this perspective.

We then summarized their collective opinion by asking "So if you are free to choose then you would like it to be done by a person, but if the need arrives then you would be okay with ready-made meals?". Everyone except for B3 agreed. B3 explained, "Even then I would prepare a simple sandwich. I wouldn't be able to eat a ready-made meal, even by looking at it I lose my appetite".

Statement 14: I prefer fraud detection to be done by humans rather than machines (algorithms).

B4 expressed a preference for the traditional approach, explaining, "Both machines and humans make mistakes". B1, on the other hand, leaned towards considering machines as potentially better. B4 then changed their mind saying "Or maybe a machine is better. If someone does not like you [or your kind] then they can put you in a lot of trouble, but a machine would not do that. Although a machine is also being made by a person, so I don't really know".

Algorithmic dashboard

The group initially agreed that it is important to prioritize avoiding wrongful accusations of fraud, as it is considered more significant than catching fraudsters by for example stating, "I think scenario A is more important, as many fraudsters with no innocent people [being accused]".

Since they all agreed, we decided to modify the scenario by asking "What if it is not so bad to be accused, let's say they only get a letter that is retracted within a month?". This caused doubt to arise among the participants.

B3 broke the silence, expressing a willingness to consider sacrificing a small number of innocent people by saying, "Maybe I would then like that scenario [where there are also innocent people accused]. But I would put my maximum on 5 people, that is not so bad in my opinion". When presented with the option of 11 innocent people being accused to catch all fraudulent people, B3 declined, deeming it too high a cost. In contrast, B2 maintained a firm stance against the idea of innocent people being wrongly accused, regardless of the number of fraudsters caught stating, "I think it is bad anyway [to accuse innocent people], no matter the number of fraudsters you catch".

The discussion highlighted differing perspectives on the balance between catching fraudsters and safeguarding innocent individuals. The participants' opinions reflected their individual values and priorities regarding justice and the protection of innocent parties.

8

Discussion and Conclusion

In this section, we will first discuss the results of our experiment and derive our findings. After that, we will assess the extent to which our research findings have effectively addressed the research questions that were defined in the introduction. Consequently, we will evaluate the achievement of our main research objective with the help of these assessments.

8.1 Discussion

In analyzing the participants' responses, several notable patterns emerged. Firstly, the participants tended to encounter more difficulty in reaching a consensus when discussing food-related topics, while they were more likely to find unity when presented with statements concerning the government or the Tax Administration. This discrepancy in agreement levels could be attributed to the participants feeling more empowered and knowledgeable in matters related to food. The opinions expressed regarding the government were also influenced by preconceived notions and biases. It is important to acknowledge that the participants struggled to understand the statements concerning the government, which may be attributed to either a lack of clarity in our phrasing or their relative unfamiliarity with governmental processes compared to their familiarity with food-related subjects.

It is worth mentioning the varying levels of participation among the participants. In the Alpha group, participant A5 was relatively quiet during the discussion, while in the Beta group, B1 was less vocal compared to the others. Interestingly, in the Alpha group, there was even a participant ('A4') who did not engage in any of the discussions.

The discussions surrounding the algorithmic dashboard proved to be particularly challenging for the participants, likely due to the high stakes involved. The complexity and potential impact of the topic contributed to the difficulty in getting a clear answer.

Nevertheless, it was encouraging to observe the nuanced perspectives demonstrated by the participants throughout the discussions. They displayed an understanding of issues such as the lack of transparency and the recognition that achieving better results often requires additional resources. Furthermore, it is worth noting that the participants' opinions were influenced by follow-up questions, indicating their willingness to reconsider and adapt their viewpoints based on new information.

Overall, the findings highlight the complexities and dynamics of participant interactions within these discussions. The differing levels of consensus, the impact of familiarity, and the influence of high stakes contribute to the rich and multifaceted nature of the

participants' perspectives. These insights provide valuable considerations for future research and discussions on these topics.

8.2 Research questions

We will first answer the research questions by analyzing the results of our experiment and our findings throughout our research.

RQ1: What is an effective way of involving marginalized communities?
We implemented a method to engage marginalized communities in the discussion by visiting local community centers in marginalized neighborhoods. In retrospect, this approach proved to be highly effective.

We believe that several factors contributed to its success. Firstly, our independent presence allowed us to establish trust and build rapport with the participants. Additionally, our shared background and familiarity with the neighborhood helped create a sense of connection and relatability. Moreover, our casual and informal approach fostered a comfortable atmosphere, enabling participants to feel at ease and actively participate in the conversation.

Based on our experience, a recommended approach would be to visit community centers in marginalized neighborhoods, accompanied by individuals who share a similar background or can establish a meaningful connection with the community members. Engaging in informal and friendly conversations with the visitors can help alleviate any apprehensions and create a comfortable environment for their active involvement.

RQ2: What are ways to make algorithms more understandable to the public?
In this study, we endeavored to address this question through our experimental approach, as outlined in subsection 5.3. Our method involved delivering an interactive presentation enriched with relatable examples and glimpses into the world of algorithmic development. This approach aimed to familiarize the participants with the concept and its significance.

We consider this experiment to be highly successful, as it resulted in a clear understanding among the participants regarding the importance of algorithms and the research we are conducting. Moreover, they demonstrated an appreciation for the nuanced impact that complex algorithms can have on modern life, beyond the commonly highlighted benefits. Consequently, we believe we have achieved our objective of making algorithms more comprehensible and accessible to the general public.

RQ3: How can we effectively incorporate the perspectives of the public?
The experiment described in subsection 5.4 serves to find out that our methodology can be seen as an effective method to incorporate the perspectives of the public. As mentioned, the input of the citizens should preferably come out as if the citizen is part of the process as either a developer or policy maker.

To this end, we feel that our chosen method, to discuss statements about food and government, was successful in determining what choices the citizens would make in certain situations. The food-related statements also worked surprisingly well in how open the participants were to share their opinion, compared to the government related questions and

even the trolley problem we discussed during the first session, where they would reluctantly give their answer.

Even though the situations are not realistic in the sense that they do not directly translate to actual algorithms, we still feel comfortable in claiming they at least give some indication as to how they would react when dealing with the actual algorithms. This overcomes barriers that would otherwise hinder the progress to get valuable input.

RQ4: How do you maintain community engagement?
The answer to this question primarily emerges from the interviews, particularly those conducted with local experts. It became evident that establishing a sense of understanding among the participants and providing tangible evidence of their contributions are crucial factors. This entails initiating the interaction with casual conversations over coffee and showcasing tangible outcomes of their involvement in the research.

Furthermore, our success with one of the groups provides another avenue for addressing this question. By approaching the topic in an engaging and enjoyable manner, while maintaining an informal and accessible atmosphere, we fostered an environment that encouraged participants to eagerly delve deeper into the subject matter.

8.3 Main research objective

Finally, we will answer whether we were successful in obtaining our research objective. To re-iterate the research objective of this thesis is to find out what effective strategies are to gain valuable input from citizens during the development life cycle of algorithms used by the public sector.

We consider our approach successful in discovering effective strategies to engage citizens in algorithmic development. These strategies included independent outreach in community centers, utilizing food as a metaphor in conjunction with government collaboration to highlight the duality, and creating a comfortable and enjoyable environment for participants to express their opinions. Through these methods, we obtained valuable insights and perspectives from the citizens involved in the process.

By answering the aforementioned research questions, we have substantiated the effectiveness of our methodology as evidenced by our evaluation. It is important to emphasize that this research serves as a foundation for future investigations, and we are satisfied with the contribution we have made in fostering further exploration in this field.

8.3.1. Scope

In this project, we considered a problem scope that was broader than common computer science research projects. Where many such projects depart from problem definitions in which particular sub-components and techniques can be addressed, our project explicitly took a holistic approach, with a lot of attention on engaging the right stakeholders, which normally would be considered out of scope for a computer scientist.

However, we believe reaching outside of the common scope is still of important for a computer scientist, as computer scientists carry part of the responsibility towards communicating between the different stakeholders.

Therefore, we feel confident in at least contributing to giving pointers to future researchers and the public sector in engaging with citizens, especially citizens that belong to marginalized communities.

9

Future work and Recommendations

Within this section, we will delve into the limitations of our research, highlight areas for potential refinement, and offer recommendations for future work.

9.1 Laying the groundwork

Our research has shown how immature the current state of affairs is on the subject of citizen participation in public algorithms. Academic and grey literature, as well as interviews, have shown that the progress of addressing these issues is still in its infancy. This is evident, for instance, in the utilization of citizen panels within the IAMA framework, as discussed in subsection 3.4.3.

Therefore, the intention of this research is to provide some basis or experiment that could be of use for a more significant stream of research. It sums up to be a foundation for communication between the public sector and marginalized citizens and bridging the massive gap that exists between these two.

9.2 Refinements

We also have refinements to the methodology we advise for further studies that build upon this.

Improve statements
While the statements we discussed during the sessions with citizens have proven valuable in capturing their perspectives, we recognize that there is room for improvement. These statements are an evolving aspect of our research and require refinement. By crafting clearer and more specific statements, we are confident we can extract more valuable insights from the participants. For instance, we can explicitly prompt participants to choose one option over the other, such as prioritizing between taste and health in statement 1. We noticed that participants often provided responses encompassing both options, which may not accurately reflect real-life dilemmas. By addressing these considerations, we aim to enhance the effectiveness and precision of our statement-based discussions with citizens.

Follow-up questions
By incorporating more effective follow-up questions, we can encourage participants to

provide a more comprehensive and well-rounded opinion. This, in turn, enables us to gain a deeper understanding of the reasoning behind their perspectives. Enhanced follow-up questions facilitate a more nuanced exploration of participants' thoughts and insights, allowing us to glean valuable insights into their perspectives and thought processes.

Interpreter
Another potential idea for future research could involve incorporating interpreters into the communication process with citizens. As discussed in subsection 6.2, one of the challenges in reaching marginalized communities is their limited proficiency in the Dutch language, which creates additional difficulties in effective communication. By introducing interpreters, this barrier can be overcome, enabling individuals from these communities to contribute their unique perspectives. The inclusion of interpreters would help facilitate meaningful dialogue and ensure that their voices are heard, ultimately enriching the research by incorporating diverse viewpoints and promoting inclusivity.

9.3 Limitations

Participants
This study is subject to limitations regarding the data collected. Due to constraints in time and resources, we were only able to gather information from a limited number of participants. Specifically, we conducted discussions with a total of 11 individuals from two community centers within a single neighborhood. It is important to acknowledge that this sample size is small and may not provide a comprehensive representation of the larger neighborhood, let alone the entire country.

Furthermore, it should be noted that the participants in this study were not diverse in terms of their backgrounds. They predominantly represented the Surinamese ethnic group within their neighborhood. As a result, the findings and perspectives obtained from this limited sample may not fully capture the diversity of experiences and viewpoints within the community.

More sessions
Other limitations due to the constraints in resources are that we were only able to conduct a limited number of sessions, which may have impacted the depth and breadth of community engagement achieved. However, despite these limitations, we made efforts to maximize the value of our research by addressing as many subjects as possible while maintaining the qualitative nature of the discussions.

With increased resources, several benefits can be realized. First, having more statements to address would allow for a broader range of topics to be explored. Second, conducting additional sessions would facilitate the incorporation of diverse experiences and enable the refinement of the quality of each session. This would provide participants with more time to reflect on their opinions and revisit previous discussions, fostering deeper engagement. Third, the availability of more resources would enable us to ask more follow-up questions and delve into topics in greater detail.

Finally, having more sessions and a longer engagement would contribute to building a stronger relationship between the moderator and participants. This would foster a more

natural and comfortable atmosphere for conversations and discussions, further enhancing community engagement.

Therefore, the allocation of more resources, particularly time, would greatly enhance the overall quality and scope of the research, allowing for a more comprehensive exploration of topics and deeper community engagement.

9.4 Future work

There remain numerous unexplored questions that, regrettably, fell beyond the scope of our research. The concept of a citizen panel for algorithms itself is still an intriguing subject that warrants further investigation. Pertinent questions arise, such as the methods of creating a fair citizen panel, or the methods to ensure diversity. Determining how diverse a citizen panel is and assessing the efficacy of diversity present other avenues for exploration.

An additional consideration might be whether it may be more advantageous to forgo the 'traditional' citizen panel and engage with focus groups or communities that are already in place. Evaluating the success of these forms of participation and understanding their concrete influence on the actual development process presents additional research questions. Experimenting with combinations of ideas, such as J1's (Table 4.1) approach combined with our approach, might also expose unseen potential.

Another crucial aspect is increased public willingness to participate. Creating the necessary incentives, the feeling that they are heard and overall interest also demand attention. Scrutinizing the potential of manipulation in these panels as well as safeguarding current stakeholders and development teams involved in the creation of governance AI from bias, also benefit from further investigation.

Looking ahead, we also envision concrete results through future endeavors. Collaborative efforts with governing bodies to co-design algorithms, as proposed in this research or in modified and advanced versions thereof, hold promise for achieving more of these concrete results.

As a concluding remark, we want to note the traction this research had as well as the support from the individuals with whom we have had the chance to interact. Everyone spoke highly of our work and was excited about what the future of this research might bring, and so are we. We hope to inspire and to have inspired others who will further develop the foundations laid in this work.

References

[1] Thomas H Cormen et al. *Introduction to algorithms*. MIT press, 2022.

[2] Natascha Just et al. "Governance by algorithms: reality construction by algorithmic selection on the Internet". In: *Media, culture & society* 39.2 (2017), pp. 238–258.

[3] Sofia C Olhede et al. "The growing ubiquity of algorithms in society: implications, impacts and innovations". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018), p. 20170364.

[4] Brent Daniel Mittelstadt et al. "The ethics of algorithms: Mapping the debate". In: *Big Data & Society* 3.2 (2016), p. 2053951716679679.

[5] Daniel Kellmereit et al. *The silent intelligence: the internet of things*. DnD Ventures, 2013.

[6] Samuel Greengard. *The internet of things*. MIT press, 2021.

[7] John Danaher et al. "Algorithmic governance: Developing a research agenda through the power of collective intelligence". In: *Big data & society* 4.2 (2017), p. 2053951717726554.

[8] T Gillespie. *The Relevance of Algorithms in: Gillespie, T. Boczkowski, PJ Foot, KA (2014) Media Technologies*. 2014.

[9] Michele Willson. "Algorithms (and the) everyday". In: *Information, Communication & Society* 20.1 (2017), pp. 137–150.

[10] Bruno Lepri et al. "Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges". In: *Philosophy & Technology* 31 (2018), pp. 611–627.

[11] Kevin C Desouza et al. "Big data in the public sector: Lessons for practitioners and scholars". In: *Administration & society* 49.7 (2017), pp. 1043–1064.

[12] Bram Klievink et al. "Big data in the public sector: Uncertainties and readiness". In: *Information systems frontiers* 19.2 (2017), pp. 267–283.

[13] Rianne Dekker et al. "Co-designing algorithms for governance: Ensuring responsible and accountable algorithmic management of refugee camp supplies". In: *Big Data & Society* 9.1 (2022), p. 20539517221087855.

[14] David Freeman Engstrom et al. "Government by algorithm: Artificial intelligence in federal administrative agencies". In: *NYU School of Law, Public Law Research Paper* 20-54 (2020).

[15] Devansh Saxena et al. "Conducting participatory design to improve algorithms in public services: Lessons and challenges". In: *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 2020, pp. 383–388.

[16] Cass R Sunstein. "Regulation in an uncertain world". In: *Speech Before the National Academy of Sciences Government-University-Industry Roundtable Meeting, Washington, DC, June*. Vol. 20. 2012.

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

[18] Philip Kraft et al. "The collective resource approach: the Scandinavian experience". In: *Scandinavian Journal of Information Systems* 6.1 (1994), p. 4.

[19] Wilhelmina Nekoto et al. "Participatory research for low-resourced machine translation: A case study in african languages". In: *arXiv preprint arXiv:2010.02353* (2020).

[20] Abeba Birhane et al. "Power to the People? Opportunities and Challenges for Participatory AI". In: *Equity and Access in Algorithms, Mechanisms, and Optimization* (2022), pp. 1–8.

[21] Brhmie Balaram et al. “Artificial Intelligence: real public engagement”. In: *RSA, London. Retrieved November* 5 (2018), p. 2018.

[22] Darin Barney et al. *The participatory condition in the digital age*. Vol. 51. U of Minnesota Press, 2016.

[23] Joy Lu et al. “Good explanation for algorithmic transparency”. In: *Available at SSRN 3503603* (2019).

[24] Solon Barocas et al. “Big data’s disparate impact”. In: *California law review* (2016), pp. 671–732.

[25] Latanya Sweeney. “Discrimination in online ad delivery”. In: *Communications of the ACM* 56.5 (2013), pp. 44–54.

[26] Danielle Keats Citron et al. “The scored society: Due process for automated predictions”. In: *Wash. L. Rev.* 89 (2014), p. 1.

[27] Frank Pasquale. *The black box society: The secret algorithms that control money and information*. Harvard University Press, 2015.

[28] Tal Z Zarsky. “Automated prediction: Perception, law, and policy”. In: *Communications of the ACM* 55.9 (2012), pp. 33–35.

[29] Tal Zarsky. “The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making”. In: *Science, Technology, & Human Values* 41.1 (2016), pp. 118–132.

[30] Madeleine Waller et al. “Why predictive algorithms are so risky for public sector bodies”. In: *Available at SSRN 3716166* (2020).

[31] Verslag-Parlementaire ondervragingscommissie Kinderopvangtoeslag. “Ongekend onrecht”. In: *Den Haag, TK* 35 (2020), p. 510.

[32] Reinier Tromp. *Fraude-algoritme op de snijtafel*. URL: https://www.vpro.nl/argos/lees/onderwerpen/algoritme/2023/fraude-algoritme-op-de-snijtafel-rotterdam-verdenkt-vooral-jonge-moeders.html.

[33] Joseph Donia et al. “Co-design and ethical artificial intelligence for health: An agenda for critical research and practice”. In: *Big Data & Society* 8.2 (2021), p. 20539517211065248.

[34] Rashida Richardson et al. “Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice”. In: *NYUL Rev. Online* 94 (2019), p. 15.

[35] Elizabeth Bondi et al. “Envisioning communities: a participatory approach towards AI for social good”. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 2021, pp. 425–436.

[36] Shintaro Miyazaki. “Algorhythmics: Understanding micro-temporality in computational cultures”. In: *Computational Culture* 2 (2012), pp. 1–16.

[37] Christopher Steiner. *Automate this: How algorithms took over our markets, our jobs, and the world*. Penguin, 2012.

[38] HLEG AI. *High-level expert group on artificial intelligence*. 2019.

[39] *Opgave ai. De nieuwe systeemtechnologie*. 2021. URL: https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie.

[40] Christian Katzenbach et al. “Algorithmic governance”. In: *Internet Policy Review* 8.4 (2019), pp. 1–18.

[41] Amir Djalali et al. “An advanced design approach to support urban transformations through multi-stakeholder collaborations.” In: *Strategic Design Research Journal* 12.2 (2019).

[42] Bogdana Rakova et al. “Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices”. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021), pp. 1–23.

[43] Rob Kitchin. “The real-time city? Big data and smart urbanism”. In: *GeoJournal* 79 (2014), pp. 1–14.

[44] Miguel Arana-Catania et al. “Citizen participation and machine learning for a better democracy”. In: *Digital Government: Research and Practice* 2.3 (2021), pp. 1–22.

[45] Sarah Giest et al. “Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective”. In: *Information Polity* 25.4 (2020), pp. 409–417.

[46] Inioluwa Deborah Raji et al. “Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing”. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 2020, pp. 33–44.

[47] Cynthia Rudin. “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. In: *Nature machine intelligence* 1.5 (2019), pp. 206–215.

[48] Zachary C Lipton. “The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.” In: *Queue* 16.3 (2018), pp. 31–57.

[49] Abeba Birhane. “Algorithmic injustice: a relational ethics approach”. In: *Patterns* 2.2 (2021), p. 100205.

[50] Stuart Hall. “Encoding/decoding∗”. In: *Culture, media, language*. Routledge, 2003, pp. 117–127.

[51] José Medina. *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and resistant imaginations*. Oxford University Press, 2012.

[52] Jennifer Pierre et al. “Getting Ourselves Together: Data-centered participatory design research & epistemic burden”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–11.

[53] Eri Bertsou. “Rethinking political distrust”. In: *European Political Science Review* 11.2 (2019), pp. 213–230.

[54] Mark Bovens et al. “The education gap in participation and its political consequences”. In: *Acta Politica* 45 (2010), pp. 393–422.

[55] Melissa S Williams. *Voice, trust, and memory: Marginalized groups and the failings of liberal representation*. Princeton University Press, 2000.

[56] URL: https://uitslagen.stembureausinrotterdam.nl/provinciale_staten23/kaart.

[57] URL: https://onderzoek010.nl/dashboard/onderzoek010/bevolking--gebied-buurt-.

[58] Lotfi El Hamidi. *Opinie: Vaccineren? Zo makkelijk laat de achterstandswijk zich niet overhalen*. Nov. 2021. URL: https://www.nrc.nl/nieuws/2021/11/19/vaccineren-zo-makkelijk-laat-de-achterstandswijk-zich-niet-overhalen-a4066118#:~:text=Vaccineren%5C%3F-,Zo%5C%20makkelijk%5C%20laat%5C%20de%5C%20achterstandswijk%5C%20zich%5C%20niet%5C%20overhalen,gevoeld%5C%2C%5C%20ziet%5C%20Lotfi%5C%20El%5C%20Hamidi..

[59] Jop van Kempen. *Experts: Geef Weerwoord aan niet-Westerse Groepen die Vaccinatie Weigeren*. Sept. 2021. URL: https://www.parool.nl/nederland/experts-geef-weerwoord-aan-niet-westerse-groepen-die-vaccinatie-weigeren~bd6c2b13/.

[60] J Bradley Cousins et al. “Framing participatory evaluation”. In: *New directions for evaluation* 1998.80 (1998), pp. 5–23.

[61] Jean A King et al. “Making sense of participatory evaluation: Framing participatory evaluation”. In: *New directions for evaluation* 2007.114 (2007), pp. 83–105.

[62] Vivianne E Baur et al. “Participation of marginalized groups in evaluation: Mission impossible?” In: *Evaluation and program planning* 33.3 (2010), pp. 238–245.

[63] Jean-Benoit Pilet et al. “Improving, Bypassing or Overcoming Representation?” In: (2021).

[64] Matthew Tenney et al. “Data-driven participation: Algorithms, cities, citizens, and corporate control”. In: *Urban Planning (ISSN: 2183-7635)* 1.2 (2016), pp. 101–113.

[65] Tahu Kukutai et al. *Indigenous data sovereignty: Toward an agenda*. ANU press, 2016.

[66] Michael McGann et al. "The rise of public sector innovation labs: experiments in design thinking for policy". In: *Policy Sciences* 51.3 (2018), pp. 249–267.

[67] Stefano De Paoli et al. "Connecting environmental action to e-participation design for young people". In: *JeDEM-eJournal of eDemocracy and Open Government* 12.2 (2020), pp. 158–191.

[68] Engin Bozdag. "Bias in algorithmic filtering and personalization". In: *Ethics and information technology* 15 (2013), pp. 209–227.

[69] Gauri Naik et al. "Will the future of knowledge work automation transform personalized medicine?" In: *Applied & translational genomics* 3.3 (2014), pp. 50–53.

[70] Aylin Caliskan et al. "Semantics derived automatically from language corpora contain human-like biases". In: *Science* 356.6334 (2017), pp. 183–186.

[71] Kate Crawford et al. "Big data and due process: Toward a framework to redress predictive privacy harms". In: *BCL Rev.* 55 (2014), p. 93.

[72] Julia Angwin et al. "Bias in criminal risk scores is mathematically inevitable, researchers say". In: *Ethics of data and analytics*. Auerbach Publications, 2016, pp. 265–267.

[73] Min Kyung Lee et al. "Participatory Algorithmic Management for Worker Well-Being". In: ().

[74] Sabine N van der Veer et al. "Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries". In: *Journal of the American Medical Informatics Association* 28.10 (2021), pp. 2128–2138.

[75] *Het systeem Fraude Signalering Voorziening (FSV)*. URL: https://www.belastingdienst.nl/wps/wcm/connect/nl/contact/content/het-systeem-fraude-signalering-voorziening-fsv.

[76] Autoriteit Persoonsgegevens et al. "Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag". In: *Onderzoeksraport z2018-2245. Den Haag: Autoriteit Persoonsgegevens. https://autoriteitpersoonsgegevens. nl/sites/default/files/atoms/files/onderzoek_belasti ngdienst_kinderopvangtoeslag. pdf* (2020).

[77] Jan. 2021. URL: https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/kamer_in_het_kort/debat-over-verslag-commissie-kinderopvangtoeslag.

[78] Jan Kleinnijenhuis. *Belastingdienst ging Vooral Achter Lage Inkomens Aan*. Nov. 2021. URL: https://www.trouw.nl/binnenland/belastingdienst-ging-vooral-achter-lage-inkomens-aan~be2da94d/.

[79] Justin-Casimir Braun et al. *Suspicion Machine Methodology*. URL: https://www.lighthousereports.com/methodology/suspicion-machine/.

[80] Saskia Klaassen et al. *Computer zegt vrouw: hoe een Rotterdams algoritme jonge, alleenstaande moeders discrimineerde*. URL: https://www.versbeton.nl/2023/03/computer-zegt-vrouw-hoe-een-rotterdams-algoritme-jonge-alleenstaande-moeders-discrimineerde/.

[81] David Davidson et al. *Zo leerde een Rotterdams fraudealgoritme kwetsbare groepen te verdenken*. URL: https://www.ftm.nl/artikelen/algoritme-gemeente-rotterdam?share=nYd%5C%2FH1Eso9Jv4vIiEq2y8JWBowRoZ9DY2A86VtQUc0RN%5C%2FiMQLhwjP1LAbD2yvGA%5C%3D.

[82] Marc Hijink et al. *Ondanks waarschuwing gingen gemeenten door met Omstreden fraudedetectiemethode*. June 2022. URL: https://www.nrc.nl/nieuws/2022/06/24/ondanks-waarschuwing-gingen-gemeenten-door-met-omstreden-fraudedetectiemethode-a4134635.

[83] *SyRI legislation in breach of European Convention on Human Rights*. URL: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx.

[84] Informatie- en Procesmanagement Team OpenData Utrecht opendata@utrecht.nl. *Algoritmeregister Utrecht*. Feb. 2023. URL: https://data.utrecht.nl/dataset/algoritmeregister-utrecht.

[85] Gemeente Utrecht. *Register verwerkingen*. URL: https://www.utrecht.nl/bestuur-en-organisatie/privacy/register-verwerkingen/.

[86] Gemeente Utrecht. *Digitale Stad*. URL: https://www.utrecht.nl/bestuur-en-organisatie/utrecht-gezonde-stad-van-en-voor-iedereen/digitale-stad/.

[87] URL: https://algoritmes.overheid.nl/nl/algoritme.

[88] Joost van Egmond. *Het bewustzijn over data-Ethiek groeit bij de overheid, maar het resultaat Moet Nog Komen*. May 2023. URL: https://www.trouw.nl/binnenland/het-bewustzijn-over-data-ethiek-groeit-bij-de-overheid-maar-het-resultaat-moet-nog-komen~b74e7539/#:~:text=Ditigale%5C%20transparantie-,Het%5C%20bewustzijn%5C%20over%5C%20data%5C%2Dethiek%5C%20groeit%5C%20bij%5C%20de%5C%20overheid%5C%2C%5C%20maar,het%5C%20resultaat%5C%20moet%5C%20nog%5C%20komen&text=Het%5C%20toeslagenschandaal%5C%2C%5C%20een%5C%20rechterlijk%5C%20verbod,toezicht%5C%20op%5C%20geautomatiseerde%5C%20beslissingen%5C%20tekortschiet..

[89] *Motie van de leden Bouchallikh en Dekker-Abdulaziz over verplichte impactassessments voorafgaand aan het inzetten van algoritmen voor evaluaties van of beslissingen over mensen #26643-835*. URL: https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2022Z06024&did=2022D12329.

[90] Apr. 2022. URL: https://www.uu.nl/nieuws/tweede-kamer-stemt-in-met-verplicht-gebruik-impact-assessment-mensenrechten-en-algoritmes.

[91] URL: https://professionals.uu.nl/en/programme/impact-assessment-mensenrechten-en-algoritmes-iama-0076f.

[92] Janneke Gerards et al. "Impact Assessment Mensenrechten en Algoritmes". In: (2021).

[93] Arvind Narayanan. "Translation tutorial: 21 fairness definitions and their politics". In: *Proc. conf. fairness accountability transp., new york, usa*. Vol. 1170. 2018, p. 3.

[94] Megha Srivastava et al. "Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 2459–2468.

[95] Alina Köchling et al. "Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development". In: *Business Research* 13.3 (2020), pp. 795–848.

[96] WH Williams. "Selection biases in fixed panel surveys". In: *Contributions to Survey Sampling and Applied Statistics*. Elsevier, 1978, pp. 89–112.

[97] Carole Paleco et al. "Inclusiveness and diversity in citizen science". In: *The science of citizen science* 261 (2021).

[98] Carsten Mann et al. "Challenging futures of citizen panels". In: *Critical issues for robust forms of public participation. A report based on interactive, anticipatory assessment of the dynamics of governance instruments* 26 (2013).

[99] *Courts Seek to Increase Jury Diversity*. URL: https://www.uscourts.gov/news/2019/05/09/courts-seek-increase-jury-diversity.

[100] Mark B Brown. "Survey article: Citizen panels and the concept of representation". In: *Journal of Political Philosophy* 14.2 (2006), p. 203.

[101] Behnam Taebi. "Bridging the gap between social acceptance and ethical acceptability". In: *Risk analysis* 37.10 (2017), pp. 1817–1827.

[102] Blake L Jones. "Effectiveness of citizen review panels". In: *Children and Youth Services Review* 26.12 (2004), pp. 1117–1127.

[103] Michael R Callaway et al. “Groupthink: Effects of cohesiveness and problem-solving procedures on group decision making”. In: *Social Behavior and Personality: an international journal* 12.2 (1984), pp. 157–164.

[104] Vetta L Sanders Thompson et al. “Strategies of community engagement in research: definitions and classifications”. In: *Translational Behavioral Medicine* 11.2 (2021), pp. 441–451.

[105] Han-Yin Huang et al. “Social Inclusion in Curated Contexts: Insights from Museum Practices”. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022, pp. 300–309.

[106] URL: `https://besmartstart.nl/`.

[107] URL: `https://denhaag.incijfers.nl/dashboard/wijkprofielen/bevolking`.

[108] Diane McDonald. “Revisiting a theory of negotiation: The utility of Markiewicz (2005) proposed six principles”. In: *Evaluation and Program Planning* 31.3 (2008), pp. 259–265.

A

Informed Consent

- Ik, Chakir el Moussaoui, doe onderzoek voor mijn afstudeerscriptie van mijn Master aan de TU Delft.
- Ik zal voor mijn onderzoek een gesprek aangaan met burgers van Nederland, om hun meningen mee te nemen en te kijken of de mening van de burger van nut kan zijn bij het ontwikkelproces van algoritmes die worden ingezet door uitvoeringsorganisaties. Hiervoor zal ik het concept van algoritmes uitleggen en gaan we zo samen dilemma's en afwegingen bespreken die gepaard gaan met het ontwikkelen van deze algoritmes met behulp van abstracte/nep algoritmes, zoals recepten. Deze gesprekken worden opgenomen, zodat ik de opinies kan gebruiken om mijn onderzoek te onderbouwen.
- Het enige aan data wat ik verzamel zijn deze opnames, waarbij u totaal anoniem blijft. Uw bijdragen worden omgezet in vertaalde citaten zodat u uw anonimiteit behoudt.
- Er zijn geen fysieke, emotionele of reputatierisico's aan verbonden aangezien het om een gesprek gaat en u compleet anoniem blijft.
- Dit is totaal vrijwillig aan uw kant en u kunt te allen tijde ervoor kiezen om u terug te trekken van dit onderzoek, dat is u goed recht.
- U kunt mij daarbij vragen om de opnames te verwijderen en dat zal ik dan ook doen.
- U krijgt mijn contactinformatie zodat u kunt aangeven als u zich wilt terugtrekken van het onderzoek, dit blijft ook geheel anoniem.
- U krijgt ook de contactinformatie van mijn supervisor indien u een klacht heeft over mij.

Mij Naam: Chakir el Moussaoui Telnr: [redacted] Mail: C.ElMoussaoui@student.tudelft.nl	Supervisor Naam: Cynthia Liem Telnr: [redacted] Mail: C.C.S.Liem@tudelft.nl

Datum:..... Handtekening:.....

B

Statements presented

1. **Ik maak mijn eten liever lekkerder dan gezond.**

☐ Eens ☐ Oneens

2. **De belastingdienst mag nog meer gegevens over ons gebruiken om beter fraude op te sporen.**

☐ Eens ☐ Oneens

3. **Ik vind het belangrijker dat mijn eten er goed uit ziet dan dat het optimaal smaakt.**

☐ Eens ☐ Oneens

4. **Ik vind het goed genoeg als de belastingdienst voornamelijk positieve resultaten laat zien.**

☐ Eens ☐ Oneens

5. **Ik vind het belangrijk om precieze hoeveelheden te gebruiken tijdens het koken/bakken.**

☐ Eens ☐ Oneens

6. **Ik vind het belangrijk dat de belastingdienst precieze gegevens gebruik van mensen (en niet een schatting).**

☐ Eens ☐ Oneens

7. **Als een recept al goed werkt is het niet nodig om wat nieuws ermee te proberen.**

☐ Eens ☐ Oneens

8. **Als een algoritme om fraude op te sporen al goed werkt hoeft het niet meer gecontroleerd/herzien te worden.**

☐ Eens ☐ Oneens

9. Ik gebruik liever wat duurdere producten om mijn eten net wat lekkerder te laten smaken.

☐ Eens ☐ Oneens

10. Ik vind het beter als de belastingdienst een methode gebruikt die langer duurt en dus ook meer geld kost om net wat betere resultaten te boeken bij fraude opsporing.

☐ Eens ☐ Oneens

11. Ik vind het niet erg dat er bedrijven zijn zoals Coca Cola die niet laten weten wat hun recept is.

☐ Eens ☐ Oneens

12. Ik vind het niet erg dat er niet volledig transparant wordt gedaan over algoritmen/technologie die worden gebruikt door de belastingdienst.

☐ Eens ☐ Oneens

13. Ik vind het niet erg als er eten gemaakt wordt door machines (kant-en-klare maaltijd) in plaats van door een mens.

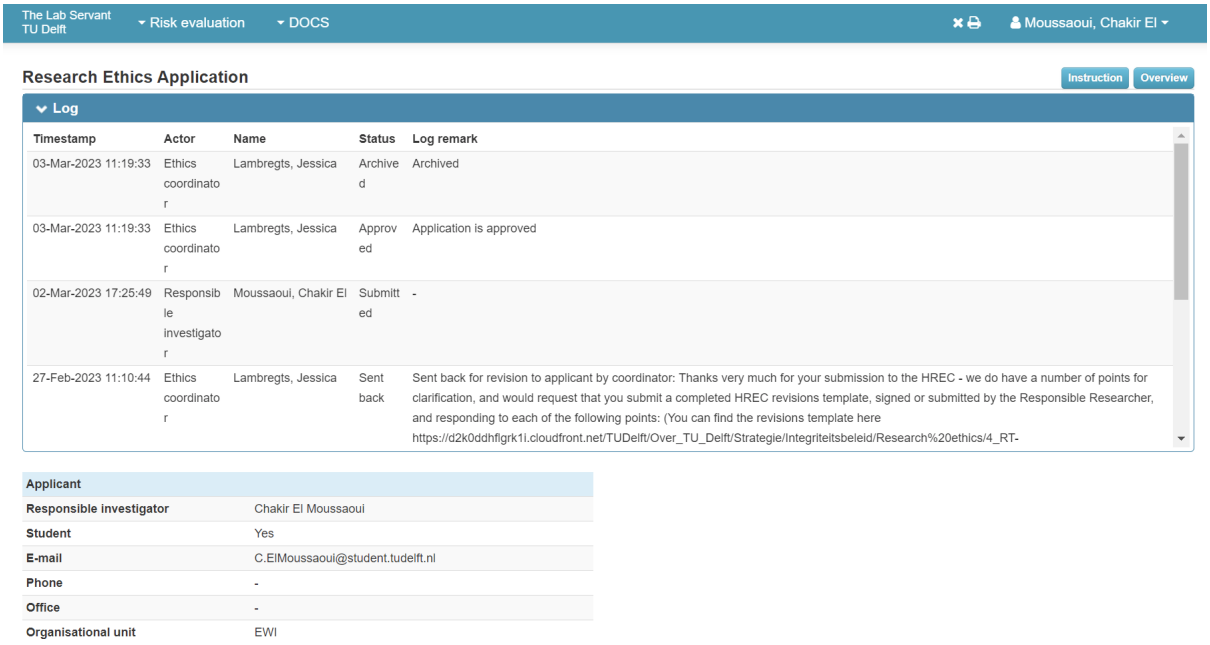
☐ Eens ☐ Oneens

14. Ik ben liever dat fraude opsporing door mensen wordt gedaan in plaats van machines (algoritmen).

☐ Eens ☐ Oneens

C

HREC application approval



The Lab Servant
TU Delft

Risk evaluation DOCS Moussaoui, Chakir El

Research Ethics Application

Instruction Overview

Log

Timestamp	Actor	Name	Status	Log remark
03-Mar-2023 11:19:33	Ethics coordinator	Lambregts, Jessica	Archived	Archived
03-Mar-2023 11:19:33	Ethics coordinator	Lambregts, Jessica	Approved	Application is approved
02-Mar-2023 17:25:49	Responsible investigator	Moussaoui, Chakir El	Submitted	-
27-Feb-2023 11:10:44	Ethics coordinator	Lambregts, Jessica	Sent back	Sent back for revision to applicant by coordinator: Thanks very much for your submission to the HREC - we do have a number of points for clarification, and would request that you submit a completed HREC revisions template, signed or submitted by the Responsible Researcher, and responding to each of the following points: (You can find the revisions template here https://d2k0ddhflgrk1i.cloudfront.net/TUDelft/Over_TU_Delft/Strategie/Integriteitsbeleid/Research%20ethics/4_RT-

Applicant	
Responsible investigator	Chakir El Moussaoui
Student	Yes
E-mail	C.ElMoussaoui@student.tudelft.nl
Phone	-
Office	-
Organisational unit	EWI

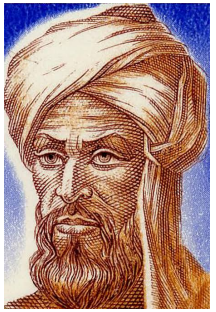
D

Slides presentation

Algoritmes

Algoritmes

- Muhammad ibn Musa al-Khwarizmi
- “Reeks instructies om vanaf een beginpunt een bepaald doel te bereiken”



Voorbeelden

- Bereidingswijze
- Instructies

Gehakt-taart

Ingredienten

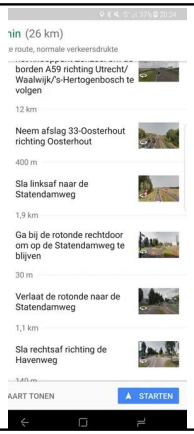
- 500 gram gehakt
- 4 teentjes knoflook
- 2 grote uien
- 2 el olijfolie
- 2 tl kaneel
- 2 tl komijn (dijnten)
- Zout en peper
- 1 tl kerrie
- 3 el gesnipperde koriander
- 3 el gesnipperde peterselie
- 10 velletjes bladerdeeg
- 50 gr boter
- 1 klein blikje tomatenpuree
- 100 gram rozijnen

Bereidingswijze

- Verwarm de oven voor op 200°C.
- Verhit de olie in een wok en fruit de gesnipperde ui en knoflook met het gehakt tot het gehakt rul is. Roer de specerijen, de tomatenpuree en de rozijnen door het gehakt en voeg een beetje water toe.
- Laat het mengsel circa 5 minuten zachtjes pruttelen tot bijna al het vocht verdampt is.
- Neem de pan van het vuur. Roer de kruiden door het gehaktmengsel en breng het pittig op smaak met zout en peper.
- Vet de ovenschaal in met boter en bekleed de schaal met de helft van het bladerdeeg en laat het deeg over de schaal hangen.
- Schep het gehaktmengsel op het deeg, vouw het overhangende deeg erover en dek het geheel af met de rest van het deeg.
- Bestrijk de bovenkant van het deeg met wat boter en bak de taart in het midden van de oven in circa 45 minuten goudbruin en gaar, dek het deeg af met aluminiumfolie als de bovenkant te bruin dreigt te worden.
- Lekker met een frisse salade

Voorbeelden

- Afslag



Voorbeelden

- Reclame
- Specifieke product

Uitdagingen

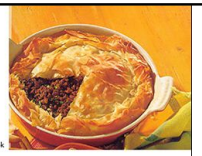
- Gegevens

Auto verzekering

Gegevens

- Ingredienten
- Hoeveelheden

Gehakt-taart



Ingredienten

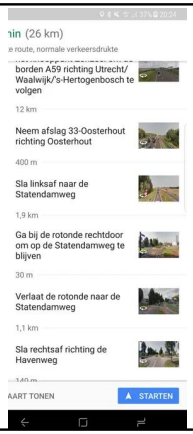
- 500 gram gehakt
- 4 teentjes knoflook
- 2 grote uien
- 2 el olijfolie
- 2 tl kaneel
- 2 tl komijn (dijnten)
- Zout en peper
- 1 tl kerrie
- 3 el gesnipperde koriander
- 3 el gesnipperde peterselie
- 10 velletjes bladerdeeg
- 50 gr boter
- 1 klein blikje tomatenpuree
- 100 gram rozijnen

Bereidingswijze

- Verwarm de oven voor op 200°C.
- Verhit de olie in een wok en fruit de gesnipperde ui en knoflook met het gehakt tot het gehakt rul is. Roer de specerijen, de tomatenpuree en de rozijnen door het gehakt en voeg een beetje water toe.
- Laat het mengsel circa 5 minuten zachtjes pruttelen tot bijna al het vocht verdampt is.
- Neem de pan van het vuur. Roer de kruiden door het gehaktmengsel en breng het pittig op smaak met zout en peper.
- Vet de ovenschaal in met boter en bekleed de schaal met de helft van het bladerdeeg en laat het deeg over de schaal hangen.
- Schep het gehaktmengsel op het deeg, vouw het overhangende deeg erover en dek het geheel af met de rest van het deeg.
- Bestrijk de bovenkant van het deeg met wat boter en bak de taart in het midden van de oven in circa 45 minuten goudbruin en gaar, dek het deeg af met aluminiumfolie als de bovenkant te bruin dreigt te worden.
- Lekker met een frisse salade

Gegevens

- Afgesloten wegen
- Actueel
- Tolwegen
- Snel of zuinig
- Auto of tram

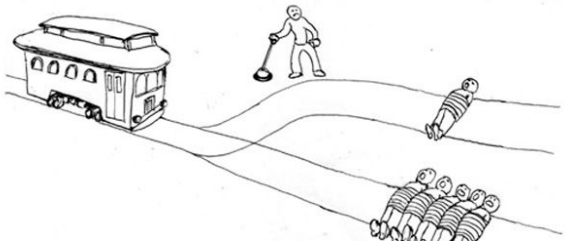


Gegevens

- Zoekgeschiedenis
- Doelgroep

Uitdagingen

- Gegevens
- Ethiek



Ethiek

- Milieuvriendelijk
- Gezond

Gehakt-taart

Ingredienten

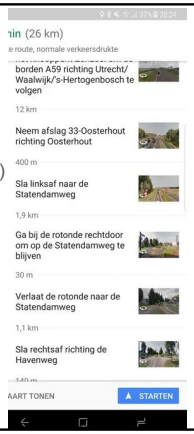
- 500 gram gehakt
- 4 teentjes knoflook
- 2 grote uien
- 2 el olijfolie
- 2 tl kaneel
- 2 tl komijn (fijnsten)
- Zout en peper
- 1 tl kerrie
- 3 el gesnipperde koriander
- 3 el gesnipperde peterselie
- 10 velletjes bladerdeeg
- 50 gr boter
- 1 klein blikje tomatenpuree
- 100 gram rozijnen

Bereidingswijze

- Verwarm de oven voor op 200˚C.
- Verhit de olie in een wok en fruit de gesnipperde ui en knoflook met het gehakt tot het gehakt rul is. Roer de specerijen, de tomatenpuree en de rozijnen door het gehakt en voeg een beetje water toe.
- Laat het mengsel circa 5 minuten zachtjes pruttelen tot bijna al het vocht verdampt is.
- Neem de pan van het vuur. Roer de kruiden door het gehaktmengsel en breng het pittig op smaak met zout en peper.
- Vet de ovenschaal in met boter en bekleed de schaal met de helft van het bladerdeeg en laat het deeg over de schaal hangen.
- Schep het gehaktmengsel op het deeg, vouw het overhangende deeg erover en dek het geheel af met de rest van het deeg.
- Bestrijk de bovenkant van het deeg met wat boter en bak de taart in het midden van de oven in circa 45 minuten goudbruin en gaar, dek het deeg af met aluminiumfolie als de bovenkant te bruin dreigt te worden.
- Lekker met een frisse salade

Ethiek

- Specifieke route
- Milieuzones
- Smalle straten druk bezocht (vrachtwagens)



Ethiek

- Privacy
- Manipulatie

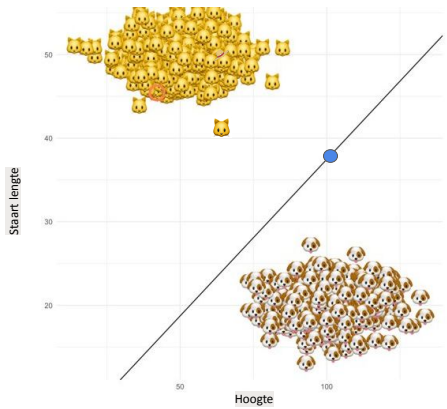
Uitdagingen

- Gegevens
- Ethiek
- Transparantie

Transparantie

- Details
- Grote lijnen
- Vertaling
- Beter?

```
// write a binary search algorithm
const binarySearch = (arr, target) => {
    let left = 0;
    let right = arr.length - 1;
    let middle = Math.floor((left + right) / 2);
    while (arr[middle] !== target && left <= right) {
        if (target < arr[middle]) {
        right = middle - 1;
        } else {
        left = middle + 1;
        }
        middle = Math.floor((left + right) / 2);
    }
    return arr[middle] === target ? middle : -1;
}
```



	🍩	🌭	🍰
Bob	☆	☆☆	☆
Jan	☆☆☆☆	☆☆☆☆	☆☆☆☆☆
Elise	☆☆☆☆	☆☆☆☆	???

	🍩	🌭	🍰
Bob	☆	☆☆	☆
Jan	☆☆☆☆	☆☆☆☆	☆☆☆☆☆
Elise	☆☆☆☆	☆☆☆☆	☆☆

Algoritmes

- Moeilijk
- Uitdagingen
- Grote invloed
- Onrechtvaardigheid
- Fouten herkennen en oplossen
- Doel

Algoritmes