Design Principles for a Non-discriminatory Financial Lending System in the Netherlands

By J. Wong

TU Delft

Deloitte.

Design Principles for a Non-discriminatory Financial Lending System in the Netherlands

Master thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

Master of Science

in Complex Systems Engineering and Management
Faculty of Technology, Policy and Management
by
Janet Wong

to be defended publicly on August 24 2023

Supervision

TU Delft
Prof dr.ir. M.F.W.H.A. (Marijn) Janssen
Dr. F. (Filippo) Santoni de Sio
MSc. R. (Reni) Sulastri

Deloitte
MSc. D. (Daan) Noordzij
MSc. M.N. (Marten) Posthumus

Preface

"Conducting this research felt like assembling a mosaic of insights, gathering pieces from diverse voices, and putting them together made me see the complexity of discrimination."
- Janet Wong (2023)

Dear reader,

I'm excited to share with you my thesis on Design Principles for a Non-discriminatory Financial Lending System in the Netherlands. This completes my MSc in Complex Systems Engineering and Management at Delft University of Technology.

The journey has been both enlightening and enjoyable. I delved deep into the world of research, learning extensively about conducting good research and the topic of discrimination in financial lending. Initially, I focused on how discrimination manifests in P2P lending. However, I quickly realized that the discrimination problem was not limited to just P2P lending. It was a broader issue that affected the entire financial lending system. Imagine my surprise when I discovered that this problem extended even further. Recently, numerous press releases, such as the controversy over DUO's student allowance, have indicated that this unequal treatment is not restricted to financial organisations but is also prevalent within governmental organizations. It seems that every day, more instances of these discriminatory practices come to light. This raised questions about why such practices continue and why they have not been addressed. I believe my research offers insights into these questions. Are you curious? Enjoy reading my thesis, and you can discover the answers for yourself.

But first, I would like to share my gratitude to the individuals who have contributed to my research. Firstly, I would like to express my thanks to Marijn, whose critical questions pushed me to elevate the quality of my research. He also made me realize that research goes beyond finding solutions for every problem that arises, as it may not always be possible, but it is a journey of exploration and understanding. Next, I am grateful to Filippo for generously sharing his knowledge of ethics, which was essential to me, given my engineering background. His insightful questions and papers assisted me to delve deeper into the subject. Next, a special thanks to Reni for introducing me to this topic. Her constant encouragement and guidance were invaluable throughout this journey, and our enjoyable conversations will be cherished. I would also like to thank Daan, whose unique perspective challenged me to explore this topic from different angles and taught me how to make my thesis more captivating for readers. Additionally, I am very thankful to Marten for meeting with me every week, providing me with that much-needed second pair of eyes, and offering helpful feedback to keep me focused and on track with my work. I extend my appreciation to the TV&A team for warmly welcoming and assisting me with my research. I am also grateful to my family and friends for their encouragement throughout this journey. Last but not least, my heartfelt thanks go to my boyfriend for his continuous and unconditional support.

As I reach the end of my student journey, I reflect with a bright smile on the incredible time I have had, the beautiful moments I have created, and the invaluable knowledge I have gained. Now, let's look forward and see what is next!

Kind regards,
Janet Wong
August 2023

Executive Summary

“Design principles against discrimination hold immense potential, but without regulatory endorsement, they gather dust on bookshelves while organizations remain inert."
- Janet Wong (2023)

Peer-to-peer (P2P) lending has emerged as a promising alternative financing option within the consumer credit market due to its ability to offer an alternative for customers who may not have access to traditional banking services. P2P lending is a form of financing that allows individuals to directly lend and borrow money from each other without the involvement of a traditional financial institution such as a bank.

There is a growing expectation that the emerging P2P lending market will provide more opportunities for disadvantaged groups through alternative financing sources. However, recent studies have uncovered instances of financial discrimination within P2P lending markets. This financial discrimination is facilitated by the adoption of new information and communication technologies (ICT) and business models, including the increased use of artificial intelligence in financial services. These algorithms have the potential to perpetuate existing prejudices and biases in lending practices, exacerbating discrimination. Moreover, human bias can contribute to these issues, as programmers may unintentionally incorporate their own biases when selecting data to train machine learning models. Therefore, addressing financial discrimination requires an approach considering technical and non-technical factors. This study adopts a socio-technical perspective to investigate this matter.

While existing research offers critical discussions, practical analyses, and technical solutions regarding discrimination, there is a lack of well-defined design principles specifically addressing discrimination challenges in financial lending in the Netherlands. This study aims to bridge the gap by providing a concrete set of design principles that architects can utilize to effectively design non-discriminatory financial lending systems, while also recognizing the potential for policymakers to draw inspiration from these principles and develop universal guidelines or other policies to address this issue more broadly.

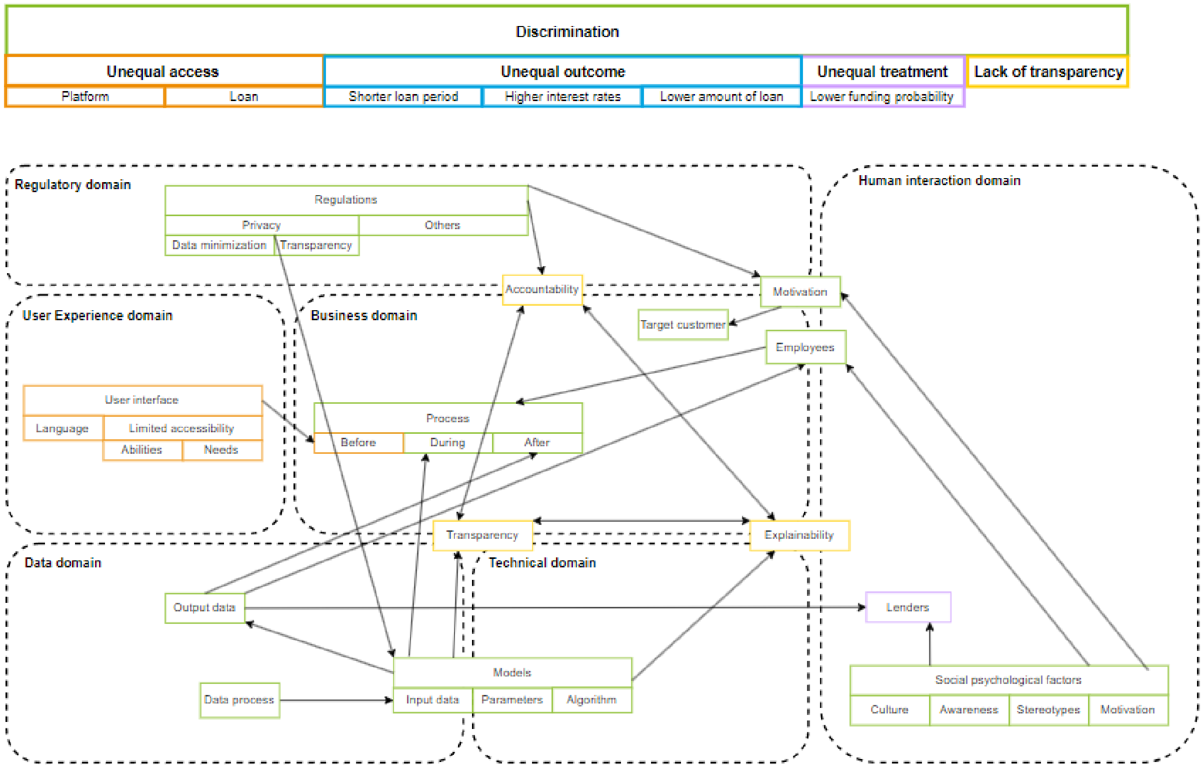
The initial research objective was to focus exclusively on the architecture of P2P lending systems. However, it became clear during the research process that addressing discrimination in P2P lending extends beyond the platform architecture. Interviews conducted revealed that discrimination is prevalent throughout the entire lending system in the Netherlands, indicating that it is a systemic issue rather than solely an architectural concern. Additionally, the research found that the highly regulated environment in the Netherlands, with capped interest rates and strict rules, has made it unattractive for P2P lending companies to operate there.

This study combined the Design Science Research approach (DSR) by Hevner (2010) and Principle Based Design (PBD) by Charosha & Janssen (2015) to derive design principles for a socio-technical topic like non-discrimination. It also introduced a layered architecture framework to examine financial organizations. This novel addition to the approach made it easier to investigate each layer separately, simplifying the process of pinpointing where in the organization discrimination might occur.

The study started with identifying the factors contributing to financial discrimination in the lending system through desk research and 13 semi-structured interviews with a wide range of scientific experts and industry professionals. To make the concept of discrimination more applicable in the context of financial lending, the study defined specific dimensions which are differential treatment, differential outcome, unequal access, and lack of transparency. This study then identified and categorized various factors that can contribute to one or more dimensions of discrimination into six domains: regulatory,

human interaction, user experience, business, data, and technical. These domains help identify the specific areas in which discrimination can arise within the lending context and how different factors may overlap across multiple domains. This highlights the complexity of discrimination, demonstrating that discrimination is not limited to one domain, and addressing a single factor can have cascading effects, either reducing or exacerbating discrimination across other domains.

The overview of the dimensions, domains, and factors can be found in the figure below:



Regulatory domain

Strict regulations, particularly regarding maximum interest rates, make it unprofitable for P2P lending platforms to operate in the Netherlands. This limitation has hindered the growth of P2P lending in the Netherlands compared to countries without such restrictions. Organizations prioritize tasks based on regulatory demands, allocating resources to urgent compliance issues. This can lead to initiatives toward non-discrimination being deprioritized. Moreover, the lack of specific guidelines and legal weight behind addressing discrimination can hinder efforts to effectively tackle discrimination in financial organizations.

Human interaction domain

Social psychological factors, including stereotypes, motivation, and culture can give rise to biases that contribute to discrimination. For instance, in the Netherlands, male-dominated industries often favour masculine traits in defining entrepreneurial success, which can disadvantage women and marginalized groups. Lack of awareness among lenders is another contributing factor. While many lenders believe they offer equal opportunities to all borrowers, closer examination often reveals biases in their investment practices. This indicates that lenders may not be fully conscious of their own biases and discriminatory behaviours.

User experience domain

Companies often prioritize user groups that generate higher profits, inadvertently overlooking the abilities and needs of smaller or fragmented user groups during the design of UI. This exclusion can lead to barriers and unequal access to loans for certain individuals. Language barriers can also lead to exclusion, as platforms that are only available in a specific language exclude non-native speakers, particularly affecting linguistic minorities.

Business Domain

Throughout the lending process, biases can emerge at various stages, making them challenging to identify and address. The reason behind this difficulty is that biases are often unconscious, and decisions are often made without clear documentation, making it difficult to trace the sources of discrimination. Moreover, the distribution of responsibilities among various teams involved in the lending process can create ambiguity regarding the decisions and actions taken by others, further contributing to the lack of transparency and accountability. Moreover, companies prioritize issues based on financial incentives and regulatory burdens, which can result in certain groups or topics being neglected if they are not considered financially lucrative or do not fall within the current regulatory focus.

Data Domain

The input data used to build credit risk models often relies on historical data collected by financial organizations. However, these often do not accurately represent the demographics of the entire population, leading to potential biases in the model's predictions. Certain types of data used in input act as proxies, introducing hidden correlations that may lead to unintended discrimination in lending decisions. Additionally, the presence of unknown unknowns in lending introduces challenges as denying a loan to an individual means never discovering whether they would have been capable of repaying it. This can inadvertently result in the exclusion of certain populations from accessing loans and opportunities.

Technology Domain

Algorithmic decision-making can lead to discriminatory outcomes, but the root cause is primarily related to humans, unrepresentative input data, and parameters rather than the algorithm itself. Algorithms are mathematical models that rely on predetermined rules and parameters, which are established by humans who may have their own biases. Biases can be introduced through the design of parameters used in these models which can unintentionally exclude certain customer groups. This can result in false negatives, where creditworthy individuals are wrongly denied loans due to the parameters considered.

Having gained an understanding of the dimensions, factors, and domains of discrimination, the study proceeded to identify the challenges encountered in realizing a non-discriminatory financial lending system in the Netherlands. This was accomplished through a combination of systematic literature review (SLR) and semi-structured interviews, which led to the identification of the following 13 challenges:

1. Lenders can discriminate based on the provided information by borrowers and the platform
2. Organisations tend to react to incidents of discrimination rather than being proactive in preventing them
3. Indirect discrimination through proxies is difficult to detect
4. Difficulty in discovering and tracing where discrimination or bias occurred throughout the process
5. Difficulty for borrowers to prove that there is discrimination
6. There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.
7. Solutions to reduce bias and discrimination can have unintended negative effects
8. Human involvement introduces biases throughout the process

9. Humans are not aware of their own biases
10. Bias can be introduced through data in many ways
11. Contextual factors can be overlooked by systems
12. Difficult to prevent all incidents due to unknown factors
13. Financial organisations tend to avoid responsibility and accountability

The aim of recognizing these challenges was to acknowledge the obstacles and difficulties associated with achieving a non-discriminatory system.

After the study unravelled the complexity of discrimination using a socio-technical understanding and the challenges associated, design principles were developed that could address those challenges in collaboration with seven business and IT architects. These principles were also evaluated in two workshops involving another group of architects. Based on the feedback received, efforts were made to improve the structure and consistency of the principles, resulting in a final set of 11 principles. To enhance clarity, the principles were rearranged and grouped into four categories: culture, governance, process, and models.

The design principles for realizing a non-discriminatory financial lending system are as follows:

Group	#	Principle Name	Statement
Culture	1	Create Awareness and Educate the Workforce	Continuously train and educate all employees to recognize and address biases
	2	Culture of Openness and Learning	Fostering an open culture where individuals are encouraged and incentivized to discuss discriminatory issues and learn from mistakes
	3	Non-Discrimination as an Objective	Set clear non-discrimination objectives and establish predefined acceptance criteria that align with these objectives
Governance	4	Dedicated Ethics Team or Officer	Establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process
	5	Have an Escalation Path for Employees and Customers and a Non-retaliation Policy	Establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy
	6	Independent Oversight	Have an external or internal independent entity that is responsible for overseeing and ensuring compliance to non-discrimination guidelines
	7	Monitor Continuously	Discrimination, both direct and indirect, should be monitored at every stage of the process
Process	8	Involve Diverse Perspectives in Collective Decision-Making	Involve diverse perspectives in a collective decision-making for important and impactful decisions to ensure that no decision is made solely by a single individual or viewpoint

	9	Traceable and Explainable Lending Process	All decisions of the lending process should be traceable and both internally understandable by employees and externally by customers and other stakeholders
Models	10	Ensure Representative and Relevant use of Data	Utilize data that is representative of the population and relevant to the lending processes
	11	Evaluate the Models and their Documentation explicitly for Discriminatory Aspects	All elements of the model (e.g., data, algorithm, parameters) used in the lending processes should be regularly evaluated and documented for potential discriminatory effects

It is important to note that there are overlaps among the principles, and their successful implementation relies on the presence of other complementary principles.

The study acknowledges that while the identified design principles can contribute to reducing discrimination in the financial lending system, they do not fully resolve the problem. This is because many discriminatory issues are deeply rooted in society. However, the principles are valuable additions to existing practices and tools, providing a framework for organizations to improve their lending practices. In addition, the involvement of regulatory authorities is essential to address discrimination in the lending system. Without regulatory incentives or oversight, organizations are not motivated to proactively implement such principles or other initiatives aimed at promoting non-discriminatory lending practices. Moreover, the design principles have the potential to be applied in other areas, such as governmental financial allowances, extending their impact beyond the financial lending system.

List of Acronyms

AI	Artificial Intelligence
AFM	Authority for Financial Markets (Dutch Financial Regulatory Body)
AP	Autoriteit Persoonsgegevens (Dutch Data Protection Authority)
BKR	Bureau Krediet Registratie (Dutch Credit Registration Office)
CCAF	Cambridge Center for Alternative Finance
CEO	Chief Executive Officer
CKI	Centraal Krediet Informatiesysteem (Central Credit Information System)
CoSEM	Complex Systems Engineering and Management
CTO	Chief Technology Officer
DNB	De Nederlandse Bank (Dutch Central Bank)
DSR	Design Science Research
EA	Enterprise Architecture
EBA	European Banking Authority
EC	European Commission
EU	European Union
GDPR	General Data Protection Regulation
ICT	Information and Communication Technologies
IS	Information System
IT	Information Technology
KPI	Key Performance Indicator
mRQ	Main Research Question
NA	Not Applicable
P2P	Peer-to-Peer
PbD	Privacy by Design
PBD	Principle Based Design
PSD2	Payment Services Directive 2
RQ	Research Question
SCORE	Subject, Context, Object, Role, and Essence layer
SLR	Systematic Literature Review
sRQ	Sub Research Question
TOGAF	The Open Group Architecture Framework
UI	User Interface
UN	United Nations
UX	User Experience
Wet AWGB	Algemene Wet Gelijke Behandeling (Dutch General Equal Treatment Act)

List of Tables

List of Figures

Table of Contents

1 Introduction

Credit plays a vital role in modern society as it allows consumers to meet their basic needs such as purchasing a home but also paying for education, or receiving medical care (Cherednychenko, 2017). In addition to benefiting consumers, credit also makes a significant contribution to the EU economy by allowing markets to operate more efficiently and stimulating economic growth (Cherednychenko, 2017). However, not all consumers have equal access to credit. Disadvantaged groups, including low-income individuals, the unemployed, migrants, disabled individuals, young people, and women, are facing difficulties in obtaining credit due to their lack of credit history or poor credit scores (Buit, 2020; Chen et al., 2020; Yum et al., 2012). This is particularly concerning as it affects those who need credit the most. The financial crisis of 2008 also made it harder for many consumers to access credit, as traditional lenders such as banks became more risk-averse and less willing to extend credit to high-risk groups (Buit, 2020). As a result, these disadvantaged consumers may turn to illegal loan sharks, who charge excessive interest rates, because of their limited access to mainstream banks (Yum et al., 2012).

In response to the abovementioned credit crunch, peer-to-peer (P2P) lending has become a promising alternative financing option in the EU consumer credit market (Buit, 2020). This is because P2P lending offers an alternative for people who may not have access to traditional banking services (Sulastri & Janssen, 2023). P2P lending is a form of financing that allows individuals to directly lend and borrow money from each other without the involvement of a traditional financial institution like a bank (Luo et al., 2011). Borrowers can request loans on an online platform without providing collateral, and lenders decide whether to grant the loan based on information provided by the borrower (Chen et al., 2020). Once a loan agreement is reached, the borrowed funds are transferred directly from the lender to the borrower, who will then repay the loan with interest to the lender over the agreed-upon timeframe. To clarify this concept, Figure 1 illustrates the differences between the interactions between lenders and borrowers in traditional banking and P2P lending systems. This form of lending has become a popular alternative for distributing credit, even during financial crises. For example, during the 2008 financial crisis, P2P lending was able to step in and provide credit to those in need, filling the void left by traditional banks (Sulastri & Janssen, 2023).

However, several studies have found that there are also inequality and discrimination within P2P lending systems. For example, Chen et al. (2020) found that female borrowers often have to offer lenders in the P2P lending market a higher rate of return to obtain a loan than male lenders. Buit (2020) found that consumers who cannot afford a bank loan also struggle to get one through the P2P lending platforms. These indicate that this promising alternative financing option cannot be regarded as inclusive either.

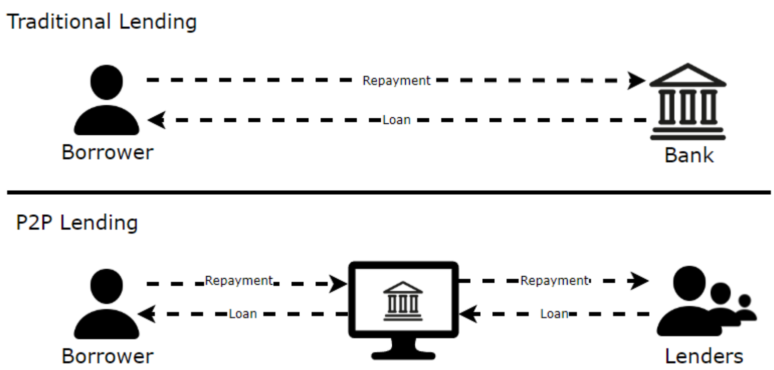


Figure 1: Difference between lenders and borrowers in traditional banking and P2P lending systems

This study aims to address the global issue of inequality, which also has been recognized as a concern by the United Nations (2018). Sulastri & Janssen (2022) remarked that various P2P lending stakeholders may have conflicting interests and concerns, including governments, IT firms, small and medium-sized enterprises, lenders, and borrowers. One example of this is the struggle P2P lending companies face in trying to achieve both profitability and access to credit for high-risk customers. These companies may prioritize bringing in as many participants as possible to generate more fees, while lenders may prefer that only those borrowers with a higher probability of loan repayment be accepted to the platform. This tension can make it difficult for P2P lending companies to align their financial objectives with those of the lenders. The presence of multiple stakeholders with differing interests adds to the difficulties in promoting inclusivity within the socio-technical P2P lending system, making it a suitable topic for a Complex Systems Engineering and Management (CoSEM) thesis.

1.1 Knowledge gap

This chapter presents the main research question (mRQ) that resulted from a literature review of scientific literature about the research problem.

This research will contribute to MSc. R. Sulastri's dissertation, which aims to design a reference architecture for P2P lending systems to enhance the financial inclusiveness of the unbanked. The study has identified a scientific gap, as where previous research primarily focused on the technical aspects of P2P lending system performance, while overlooking inclusiveness and access to credit facilities for micro-small-medium enterprises. Additionally, most studies have emphasized financial profitability rather than addressing societal and governmental needs for inclusiveness. There are also currently no socio-technical guidelines that address both technical and social issues.

1.1.1 Financial discrimination

To increase the feasibility of the research, the scope of this study is limited to financial discrimination within the broader concept of financial inclusivity. This study focus on the legal definition of discrimination stating that all individuals should be treated equally under equal circumstances, and discrimination based on protected characteristics is prohibited. This focus is motivated by extensive literature confirming the presence of financial discrimination against females, ethnic minorities, immigrants, and individuals with limited education in traditional banking systems (Wu et al., 2023). According to their definition, financial discrimination refers to the economic constraints vulnerable social groups face due to their personal characteristics, including gender, race, religion, and educational background. Gajane & Pechenizkiy (2017) have a similar definition and defined discrimination as the 'unfavourable treatment of people due to the membership to certain demographic groups that are distinguished by the attributes protected by law' (p. 1). There is a growing expectation that the emerging P2P lending market will provide more opportunities for disadvantaged groups through alternative financing sources. However, recent studies revealed that there are also financial discrimination in P2P lending markets (Wu et al., 2023). This financial discrimination introduces new forms of discrimination enabled by new information and communication technology (ICT) and business models, such as the growing use of machine learning and artificial intelligence (AI) in financial services (Kozodoi et al., 2022; Wachter et al., 2020). Machine learning algorithms have for example the potential to improve financial procedures, including loan approvals and lending decisions (Kozodoi et al., 2022). However, these algorithms often reinforce existing prejudices and biases in lending practices, worsening discrimination (Lee & Floridi, 2021; Wachter et al., 2020). Moreover, Wang et al. (2020) found that lenders using algorithms for decision-making, along with human input, displayed a tendency towards regional discrimination. This supports Chander's (2016) theory that a seemingly transparent and neutral algorithm may still generate discriminatory outcomes, potentially exacerbating socioeconomic polarization.

Despite the existence of laws prohibiting lending discrimination, the growing use of machine learning algorithms in the credit market has given rise to new challenges (Kumar et al., 2022). These algorithms can lead to various forms of unfairness, raising concerns about the potential for illegal discrimination by financial institutions. Moreover, human bias can contribute to these issues, as programmers may unconsciously include their own biases when selecting data to train machine learning models (Chander, 2016). Therefore, this implies that addressing financial discrimination requires an approach that considers both technical and non-technical factors. Regulators are also confronted with the task of updating their enforcement methods and offering new guidance accordingly (Kumar et al., 2022). However, according to Kumar et al. (2022), discussions around algorithmic discrimination often remain abstract and not grounded in actual practices, data, and algorithms used in the sector. It is important to recognize that studies like Noble (2020) and Eubanks (2018) do offer more detailed and practical analyses of the problem. Nevertheless, they do not offer concrete design guidelines or solutions to address the issue at hand. This highlights the importance of transforming these critical discussions, concerns, and challenges into tangible guidance, such as principles that prevent discrimination and ultimately support inclusivity.

Currently, the literature does not provide clear guidance on how to safeguard inclusivity, specifically non-discrimination, when designing and setting up information systems and architecture. However, established principles for other values, such as privacy, can serve as valuable lessons for addressing financial discrimination. Privacy by Design (PbD) is one such example, comprising seven principles and offering a proactive approach to embedding privacy into the early stages of ICT design (Alshammari & Simpson, 2017)

The concept of PbD originated in the 1970s and 1980s as a response to growing concerns about personal data collection and potential government surveillance (Bennett, 2018). As a result, experts established study commissions to develop a consensus on how to address this privacy problem through public policy. This led to the development of data protection and information privacy statutes based on a similar set of principles that have spread around the world in several stages (Bennett, 2018). Cavoukian advocated for a proactive approach to privacy protection, which emphasized the importance of incorporating privacy into ICT design and architecture from the beginning, resulting in PbD (Cavoukian et al., 2010). This approach gained international recognition in the 2000s and became a foundational principle of the EU's General Data Protection Regulation (GDPR) in 2016, requiring organizations to implement PbD.

With the increasing challenges of discrimination, a situation like the development of privacy concerns may arise. However, the study of algorithmic discrimination is still in its early stages, without any established principles to address it. This study aims to develop a set of principles to prevent discrimination in financial lending systems. This could serve as an initial step towards creating universal principles to prevent discrimination in information systems and architecture.

1.1.2 Design principles

This section dives deeper into what is already researched on design principles. This study offers guidance on how to design an inclusive P2P lending architecture contributing to MSc. R. Sulastri's dissertation. While there are various methods for providing guidance for the design of a (reference) architecture, this study has deliberately chosen to focus on design principles due to their important role in guiding complex socio-technical problems that require the integration of both technical and social aspects into the design process (Bharosa & Janssen, 2015; Clegg, 2000). Although there is no uniform definition for design principles, Bharosa & Janssen (2015) have noted that most studies define principles as having a normative or prescriptive nature that offers guidance for information system (IS) design. This study will adopt this definition.

To further explore the knowledge gap regarding design principles to prevent discrimination in architecture, a literature search was conducted in March 2023 on Scopus. Several studies were found that extracted design principles for IS, such as Seidel et al. (2018) who extracted design principles for IS supporting sensemaking in environmental sustainability transformations, and Gimpel et al. (2022) who extracted design principles for mobile eco-driving feedback IS. Moreover, the Handbook of van den Hoven et al. (2015) proposes an overview of many different ethical and societal values, and how they can be embedded in design through Value-sensitive Design. In addition, there are also studies that have identified design principles that aim to safeguard specific values in the design of IS. For instance, Matheus et al. (2021) focused on designing IS that promote digital transparency in government, Kosmas et al. (2020) focused on safeguarding privacy in social computing systems, and Bharosha and Janssen (2015) focused on safeguarding information quality for emergency management processes.

While there are numerous studies on principles related to inclusivity, almost all of them are closely linked to universal design. For instance, the work done by Hughes et al. (2020) leverages universal design principles to develop situational awareness IS that consider the specific needs of individuals with situational disabilities during disaster situations. Although universal design principles have been identified for inclusivity, they are mainly used in other domains such as education (Burgstahler & Cory, 2010; Cumming & Rose, 2022; Pittman & Heiselt, 2014; Wijeratne et al., 2022) or physical architecture (Abdelmaksoud, 2022; Lytvynenko et al., 2020; Mosca & Capolongo, 2020; Soydaş Çakır & Tosun, 2022). There are no research that has evaluated whether these universal design principles are effective in the context of P2P lending or in addressing challenges presented by new ICT, such as the biases in machine learning algorithms. It is also found that the applicability of these principles to specific contexts are also limited. That is why many researchers have developed additional design principles to complement or adjust these existing principles to better suit their own subject areas (Gjøsaeter et al., 2021; Montanha et al., 2008; Yiangkamolsing et al., 2010).

No research was found that focuses on extracting design principles to prevent discrimination in financial lending systems in general or architectures. Conversations with various architects within Deloitte Consulting Netherlands also confirmed this. These principles would ensure that non-discrimination is embedded in the organization's design from the start and across all layers and are therefore important for society. By following these principles, companies not only comply with existing discrimination regulations, but also prepare themselves for future regulations in this area. This shows parallels to what has been observed with privacy. These principles can also be used by policy makers as a starting point or inspiration for developing concrete principles for companies to proactively prevent potential discrimination issues from the beginning of its design. It is also important to note that there are policies regarding AI ethics that indirectly touch upon discrimination, such as the Ethics Guidelines for Trustworthy AI, the EGE Group Report on AI Robotics and Autonomous Systems. However, these documents provide high-level guidelines that are broad, not directly applicable, and can be perceived as ambiguous, leading to potential varied interpretations.

To summarize, discrimination could undergo a similar path as privacy of having concerns, developing principles, and eventually incorporating them into universal principles integrated in regulations. However, currently there is a gap in translating non-discrimination and its challenges into principles that can be used as a basis for developing universal principles.

1.2 The scope of this research

Considering this study's time constraints, it is necessary to narrow the scope. Resulting in a focus on the specific dimension of discrimination instead of the broad concept of inclusivity. The next section will provide more details on this concept. It is important to note that there are currently no established principles for integrating non-discrimination into design practices, as confirmed by various (enterprise) architects. Initially, the research aimed to focus solely on the architecture of P2P lending systems. However, it became clear during the research process that addressing discrimination in P2P lending extends beyond the platform architecture. Interviews conducted as part of this study revealed that discrimination is prevalent throughout the entire lending system in the Netherlands, indicating that it is a systemic issue rather than solely an architectural concern. The decision to study the Netherlands specifically was twofold: firstly, the EU has recently been actively addressing concerns on ethical (AI) design, making a case study in the Netherlands, as part of the EU, highly relevant. Secondly, the Netherlands have a highly diverse population in terms of ethnicity and culture. By examining discrimination in financial lending in the Netherlands, insights can be gained into the challenges faced by minority groups in accessing loans, thereby understanding the barriers they might encounter Consequently, the mRQ has been revised to include not only the P2P lending architecture but also the broader lending system in which P2P lending operates. Moreover, this study focuses on the actions that financial organizations can take to improve non-discrimination in lending. It does not address factors beyond their control. This means that the principles are intended to guide financial organizations in addressing discrimination within their own processes and practices.

1.3 Main research questions

Given these findings, the identified research gap and the scope, the following mRQ is formulated:

What are design principles for a non-discriminatory financial lending system?

2 Theoretical background

This chapter aims to provide a theoretical foundation for the following chapters in this report and is divided into three main sections. The first section explains P2P lending. The second section presents a conceptual architecture framework that will be used in the study. Lastly, the final section provides background information on the definition and types of discrimination.

2.1 Peer-to-peer Lending System

This section explains the concept of P2P lending, the lending process and the key stakeholders involved within the Dutch context.

2.1.1 What is P2P Lending

The rise of digitization and internet technologies has transformed the way people connect with each other (Abbassi & Iyer, 2021). Many industries have therefore undergone huge changes in the way they do business. This also includes the financial lending market which has undergone a transformative process with the emergence of a new version of the traditional private lending market in the form of online credit markets (Abbassi & Iyer, 2021). One type of lending market are P2P lending markets, where private lenders and borrowers can directly engage with each other without needing a financial institution as an intermediary. The online model of private lending only requires a platform that facilitates search and matching to connect borrowers with lenders. The decision of whether a loan is created, is given directly to the lenders and borrowers themselves. Borrowers provide details about the purpose of their loan request and their current financial situation, such as income or existing credit lines. Lenders, in turn, can offer loans with interest rates based on this information (Bachmann et al., 2011). However, Abbassi & Iyer (2021) also highlighted that many platforms now require lenders to auto-select loans, meaning they no longer have the option to choose to which borrower to lend. For borrowers, P2P lending presents an alternative to traditional banking. It allows them to obtain loans without involving financial institutions in the decision-making process. It also has the potential to provide better loan conditions compared to traditional banking systems. On the other hand, lenders view P2P lending as an investment opportunity in which the risk is linked to borrower's credit ratings (Bachmann et al., 2011).

An increasing amount of scientific research on P2P lending in recent years indicates the growing interest and the importance of this relatively new field. In 2005, Zopa in UK established the first lending platform that provided a forum for private lenders and borrowers. However, it was Prosper.com that significantly contributed to the wave of scientific interest by releasing the platform's entire dataset to the public in 2007 (Bachmann et al., 2011). Since then, the number of platforms has increased significantly, with 52 distinct business models worldwide by the end of 2018 (Rau, 2018). The Cambridge Center for Alternative Finance (CCAF) identifies over 3,500 platforms in 182 countries (Rau, 2018). These platforms generate revenue by imposing certain fees on the intermediation of the transaction. They are attractive alternatives for small borrowers because of the absence of collateral. Without these platforms, borrowers would have to rely on payday lenders or high interest credit cards (Adams et al., 2009; Iyer et al., 2016). However, it is important to note that platforms can use different ways to determine the interest rate of a loan, which will be described in the next sub-section.

2.1.2 The main processes of P2P lending

The general P2P lending process is almost the same across different online P2P lending platforms. Users, both borrowers and lenders, register by providing personal information and undergo a credit rating process specific to the platform. Borrowers initiate the lending procedure by specifying the desired loan amount, maximum interest rate, and additional details. Lenders either directly choose a borrower or contribute to a fund pool managed by the P2P lending company. When a borrower's request is fully funded, the transactions are reviewed by the lending intermediary, and additional credibility documents may be requested. Once approved, funds are transferred from lenders' accounts to borrowers' accounts (Wang et al., 2015).

Different methods are used to determine the final interest rate for the borrower. Proper.com for example uses an auction process. Borrowers indicate their maximum interest rate that they are willing to pay. During a specific timeframe (e.g., 14 days on Prosper), lenders can place bids with their respective minimum interest rate and amount. Even after the loan has been fully funded, lenders can continue to place bids with better terms to compete with other lenders. If more bids are received than funds required, the bids with the lowest minimum interest rates are chosen. All lenders whose bids are accepted will receive the interest rate of the highest bid included, independent of the smaller minimum interest rates specified in their own bids (Bachmann et al., 2011). Other platforms, like Smava.de from Germany, have a different approach to determine the interest rate. Instead of relying on an auction process, these platforms calculate interest rates based on the borrower's financial and demographic characteristics. Once the loan request is fully funded, the bidding process concludes as additional bids would not impact the resulting interest rate (Bachmann et al., 2011).

P2P lending has a more complex process for determining interest rates compared to banks. Unlike banks, where rates are fixed, P2P lending allows borrowers and lenders to negotiate and agree on the interest rate. This flexibility could enable a fairer rate that meet the specific requirements of individuals and the market. On the other hand, banks can be more complex in other parts of the process, like managing loans after they have been approved, as they follow more complex standardized procedures (Wang et al., 2015).

2.1.3 The stakeholders of P2P Lending

There are multiple stakeholders involved in P2P lending systems, each with its own role (Figure 2). Table 1 presents a concise overview of these stakeholders, along with their respective roles and interests. The roles and interests of P2P lending platforms, borrowers, and lenders have already been addressed in the previous sections. As for the regulatory bodies, their roles and interests are derived from their official website. This overview forms a basis for understanding the context of this study.

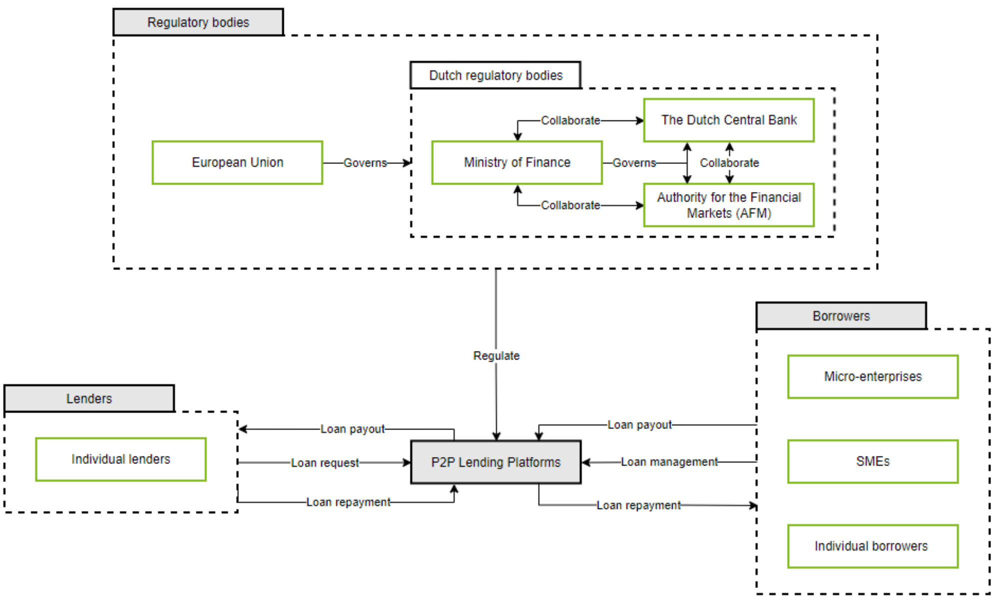


Figure 2: Stakeholder landscape of P2P lending systems

Table 1: Description of stakeholders in P2P lending systems

Stakeholder	Description	Role	Interests
P2P Lending Platforms	Online intermediaries connecting borrowers and lenders	Manage transactions, assess risks, set loan conditions	Generate revenue
Lenders	Surplus-funds individuals or businesses	Provide funds, choose borrowers	Maximize returns and minimize defaults
Borrowers	Funding-seeking individuals or businesses	Apply for and repay loans	Quick access to funds, reasonable interest rates, easy process
Ministry of Finance	Dutch government's financial body	Regulate (P2P) lending market, enact financial laws	Maintain economic stability
Dutch Central Bank (DNB)	The Netherlands' central bank	Supervise and monitor P2P platforms	Maintain stability of the financial system
The Authority for the Financial Markets (AFM)	Dutch financial market regulator	Oversee the operations of P2P platforms	Promote fair, transparent financial markets, prevent fraud.
European Union (EU)	Union of 27 European countries	Providing EU-wide regulations	Maintain a stable, competitive EU economy

2.2 Architecture framework

Since financial organisations consist of both technical and organizational components, an Enterprise Architecture (EA) framework is selected, as it is a holistic strategy that aligns an organization's business with its information technology (IT) to achieve objectives. Several EA frameworks are available, including The Open Group Architecture Framework (TOGAF), Zachman EA Framework, and EA3Cube (Nasef & Bakar, 2020), but this study has chosen to adopt a layered architecture framework.

This study combines TOGAF with the Deloitte Horizontal Framework to create a comprehensive framework for structuring the thesis findings (Figure 3)

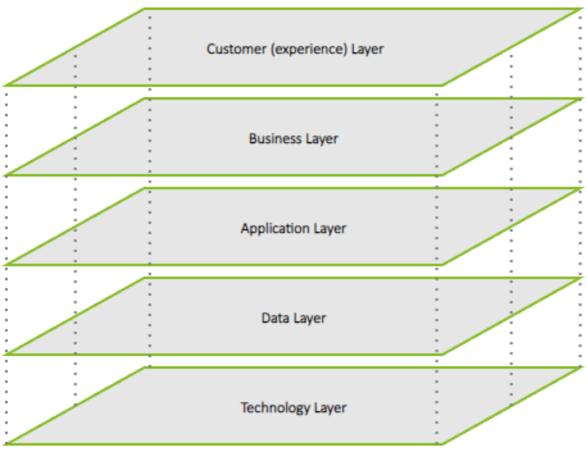


Figure 3: Layered Architecture Framework

Various layered frameworks are available. Each framework has its unique set of layers. As illustrated in Figure 3, the framework developed for this study consist of the following layers: customer (experience) layer, business layer, application layer, data layer, and technology layer. The last four layers are derived from TOGAF. Furthermore, Mohamed et al. (2020) have noted that the majority of EA frameworks incorporate these same four layers.

TOGAF is an open-source framework that was developed in 1995 by The Open Group to provide organizations with the tools and methods needed to build EAs (Nasef & Bakar, 2020). One of the strengths of TOGAF is its cost-efficiency, as it uses a simplified approach to design, plan, acquire, and integrate IT architecture to the business. It is also flexible, adaptable, and more practical to use than other existing framework solutions. Additionally, TOGAF is an open framework that is freely accessible to anyone and supported by a large community, which enhances its credibility (Nasef & Bakar, 2020). This is useful in this study as many architects are already familiar with TOGAF, which means that less time is needed to explain the framework during interviews.

Deloitte has developed its architecture framework known as the Horizon architecture. This framework is based on years of experience designing and implementing architectures for clients (Clark et al., 2023). They discovered that architectures are often misunderstood and undervalued, leading them to develop a framework that enables organizations to win and change effectively in the marketplace. As Clark et al. (2023) define, the Horizon architecture is a "technology architecture that exhibits a thoughtful, intentional, and value-oriented configuration across the different layers of architecture" (p. 9). The framework consists of 5 layers namely customer and user experience (UX), business architecture, application and integration, data and analytics, and infrastructure.

One notable difference between TOGAF and Horizon architecture is that the latter includes an additional layer called the customer (experience) layer, which relates to the UX, such as UI/UX design. This additional layer is incorporated into this study as scholars find that platforms are often considered non-inclusive due to their interface, excluding users who do not speak the language the platform is in (Ferati et al., 2016; Mack et al., 2020). By incorporating this layer, it is assumed that relevant inclusivity challenges that hamper the inclusivity of P2P lending architecture can be identified. The horizontal architecture framework also consists of all the layers of TOGAF. However, horizontal layers' terms are changed to those used in TOGAF to make it more widely recognized and easier to use for architects. For instance, the lowest layer in Horizon architecture is called infrastructure, in this thesis the TOGAF's term technology layer is used instead.

2.2.1 Advantage of layered framework

Several scholars have noted the overall benefits of the layered model and its usefulness in deriving design principle. For example, Lange et al. (2012) argued that the specification of EA principles for each individual layer of the EA is important. They found that extending EA principles from the business layer to the infrastructure layer can address and solve a wide range of architectural challenges. Chakraborty & Sarkar (2017) proposed the SCORE (Subject, Context, Object, Role, and Essence layer) framework for EA, which comprises five loosely coupled layers, allowing any lower-level layer to be changed according to business requirements without altering the upper-level layers. This feature increases reusability and enables other applications to use the functionalities exposed by the layers. Karakaxas et al. (2000) stated that layered layers can be used to separate architectural concerns or viewpoints. This makes the architecture easier to understand and managed.

In addition to the overall benefits of using a layered architecture, there are also benefits specific to this study. Firstly, this framework helps in the targeted and structured data gathering by identifying suitable stakeholders for each layer. For example, a solution architect can be interviewed for the technology layer, and an information architect for the data layer. This method helps to collect data that is more focused and relevant to this study. Additionally, conducting targeted interviews provide a more comprehensive understanding of the challenges faced by each layer, and potential solutions. By interviewing individuals with specialized knowledge and expertise, this study could reveal insights that may be missed in a broader data gathering process involving a general group of individuals. The approach will be discussed in more details in the following Chapter.

2.3 Discrimination

In the Netherlands, despite not having specific regulations in place for discrimination in P2P lending, there are general discrimination laws and regulations in place for various contexts, including the financial sector. This study focuses on the main principles of the law, providing an overview while leaving out specific details. As shown in the stakeholder overview (Figure 2) Dutch regulatory bodies are governed by the EU. The EU has several directives against discrimination that must be implemented by member states, including the Netherlands. Dutch law has incorporated the equality and non-discrimination directives. The Netherlands primarily implemented these directives through Article 1 of the Dutch Constitution, which states that all people in the Netherlands are equally treated in equal circumstances, and discrimination on the grounds of religion, belief, political opinion, race, gender, or any other grounds is not permitted. While Article 1 does not provide detailed rules on protected characteristics, EU discrimination directives are transposed into Dutch law through more comprehensive and specialized legislation. Appendix A provides a more detailed legal analysis of this matter, resulting in a list of 11 distinct variables that are protected from discrimination as outlined in regulatory documents. Based on this, from the perspective of regulatory bodies in the context of P2P lending, discrimination can be defined as follows:

All individuals (borrowers and lenders) should be treated equally under equal circumstances, and discrimination based on any of the following 11 protected characteristics according to the European and Dutch law is prohibited: belief, gender, civil status, nationality, political opinion, race, religion, sexual orientation, age, disability, and chronic illness.

Empirical evidence shows that discrimination continues to persist, despite the regulations and laws in place to prevent it. It is particularly prevalent in the financial sector. Wu et al. (2023) show this widespread issue of financial discrimination. They found that literature confirms the existence of financial discrimination against women, ethnic minorities, immigrants, and individuals with low levels of education. According to their definition, financial discrimination refers to the economic constraints faced by vulnerable social groups due to their personal characteristics. These characteristics include gender, race, religion, and educational background. An example of an economic constraint is higher interest rate (Wu et al., 2023). Much research has already shown existence of such financial discrimination in the traditional banking system (Atkins et al., 2022). Only more recently have researchers started to study discrimination in P2P lending (Wu et al., 2023). This will be discussed in more details later in this study.

2.3.1 Different types of discrimination

Two types of discrimination exist from a legal perspective, direct and indirect (systematic) discrimination (Daniel et al., 2014). Zuiderveen Borgesius (2020) mentions that these are referred to as disparate treatment and disparate impact. Direct discrimination or disparate treatment refers to rules, laws, or events that explicitly target minority or disadvantaged groups based on specific attributes related to their group membership. There is an intention to discriminate. On the other hand, there is indirect discrimination or disparate impact which involves rules or procedures that do not explicitly target or mention discriminatory attributes but can still lead to unfair treatment. Without having any discriminatory intentions, these might result in discriminatory outcomes and are therefore classified as indirect discrimination (Daniel et al., 2014).

In addition to the legal view, there is also the economic view on discrimination where discrimination can be divided into “taste-based” or “statistical” discrimination (Morse & Pence, 2021a). Taste-based discrimination occurs when decision-makers have biases and willingly sacrifice other priorities, like hiring the most qualified workers, to satisfy those biases (Morse & Pence, 2021a). Human biases refer to judgment, decision making or perception that are linked to subjective experiences, beliefs, preferences, or stereotypes. People rely on simple intuitive heuristics or rules of thumb in making their judgments often producing biases and systematic errors (Kruglanski & Ajzen, 1983). Statistical discrimination, on the other hand, happens when decision-makers lack certain information and use group averages or correlated variables to make judgments. This can be illegal if it involves the aforementioned protected categories, but decision-makers may use other factors to estimate missing information. Unlike taste-based discrimination, statistical discrimination is often driven by profit maximization rather than personal biases (Morse & Pence, 2021a).

In addition, Morse et al. (2021) highlighted the distinction made between 'overt evidence' and 'comparative evidence' of dissimilar treatment. The former refers to situations where a lender openly discriminates on a prohibited basis, while the latter refers to cases where a borrower is treated differently based on factors strongly related to a prohibited basis. For example, using grey hair as a proxy for age in decision-making can be considered comparative evidence of age discrimination. The laws of the EU and Netherlands also cover indirect discrimination, which occurs when a policy or practice that appears neutral actually puts individuals of certain protected characteristics at a disadvantage compared to others. To be considered lawful, such provisions needs to be justified by a legitimate aim and the means of achieving that aim must be appropriate and necessary.

3 Research design

This chapter presents a structured approach to the mRQ by breaking it down into sub research questions (sRQ). It includes details on data requirements, data collection methods, analysis tools, and a flow diagram to demonstrate the relationship between the sRQ and the mRQ. Figure 4 shows an overview and simplified version of the research design.

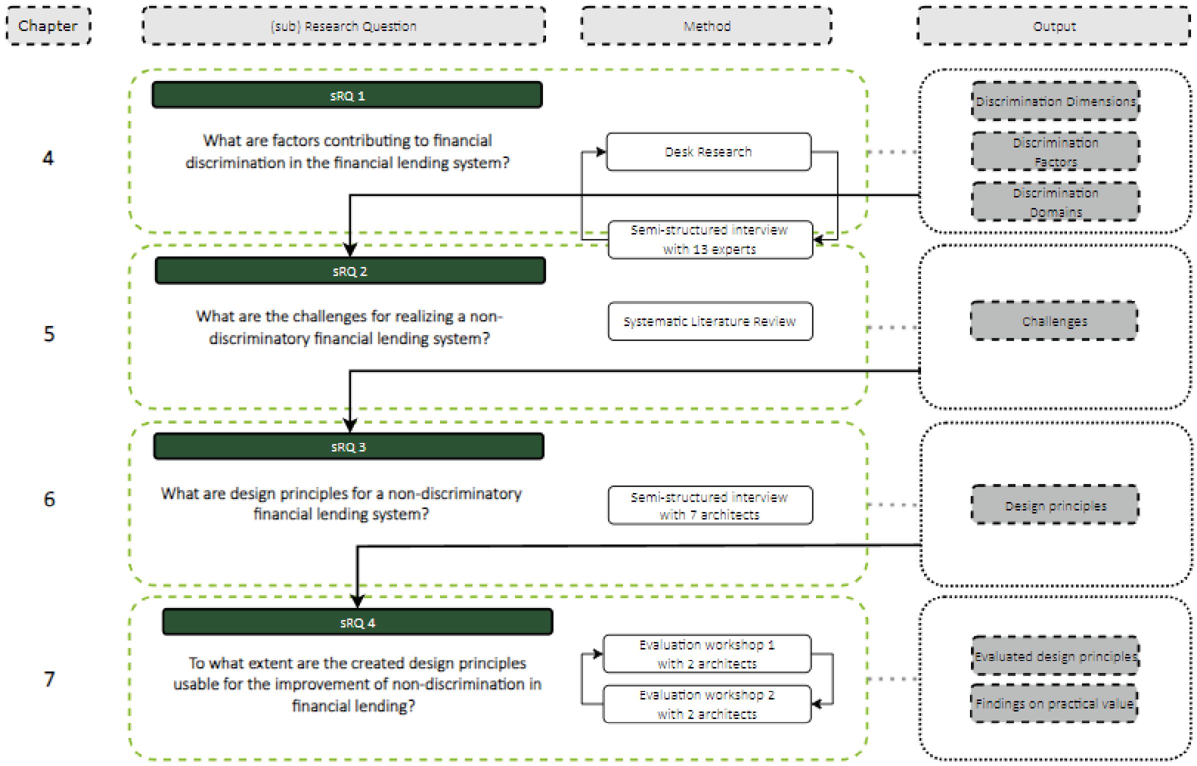


Figure 4: Overview and simplified version of the research design

3.1 Research approach

This section discusses the research approach of the study. The objective of the study is to discover what the design principles for a non-discriminatory financial lending system are. This can be classified as design-oriented research as the outcome of the study will serve to address discrimination challenges. The Design Science Research (DSR) approach as proposed by Hevner (2010) has been chosen (Figure 5) and has been adapted to the context of IS and considered best practices from design research traditions of other fields (Hevner et al., 2008). It involves integrating (academic) research and environmental knowledge into the design and evaluation of IT artifacts, with the goal of advancing existing research knowledge. It is important to note that there is no clear definition of what IT artifacts within DSR are. However, Dresch et al. (2015) stated that it can take various forms such as models, methods, frameworks, or design propositions. Principles can be part of DSR (Bharosa & Janssen, 2015). Other studies have also utilized the DSR approach to design principles such as Corten (2018), Seidel et al. (2018) and Matheus et al. (2021). This method is suitable for this research, as it emphasizes the creation of IT artefacts as a solution to solve a complex problem using an interdisciplinary approach. This is relevant as the discrimination problem in financial lending systems can be considered a complex multi-actor issue, as discussed in Chapter 1.

This study utilized the Principle Based Design (PBD) proposed by Bharosa & Janssen (2015) as their PBD approach enables the systematic extraction of principles that can be used to design IS. Moreover, Bharosa & Janssen (2015) argued that PBD is focused on achieving goals rather than adhering to rules, and since architects are given freedom to implement principles, it is expected that there will be greater commitment and less resistance in situations with multiple actors. PBD prioritizes doing what is right in a manner that architects deem most appropriate, given the circumstances. Therefore, design principles are often more agreeable to stakeholders because they are abstract and less restrictive than requirements and constraints. Additionally, agreeing on principles still allows architects and developers the freedom to customize systems to meet local circumstances and needs. This approach is therefore particularly useful in the dynamic P2P and financial lending context, which involves multiple stakeholders (e.g., borrowers, lenders, and platform owners) with different needs. Moreover, it is well-suited for scenarios that involve high levels of uncertainty, where the different aspects of a complex problem cannot be fully predicted or specified in advance (Bharosa & Janssen, 2015). The focus of this study is solely on developing principles for creating non-discriminatory financial lending system, without considering the trade-off between different values. Therefore, architects are responsible to balance these principles with other values, such as profitability and transparency, to create a viable business case.

Thus, this study aims to establish principles for non-discrimination, without exploring any potential interactions between these principles and other design principles. Consequently, it is uncertain how these principles will translate into practical applications which will be discussed in the limitation section.

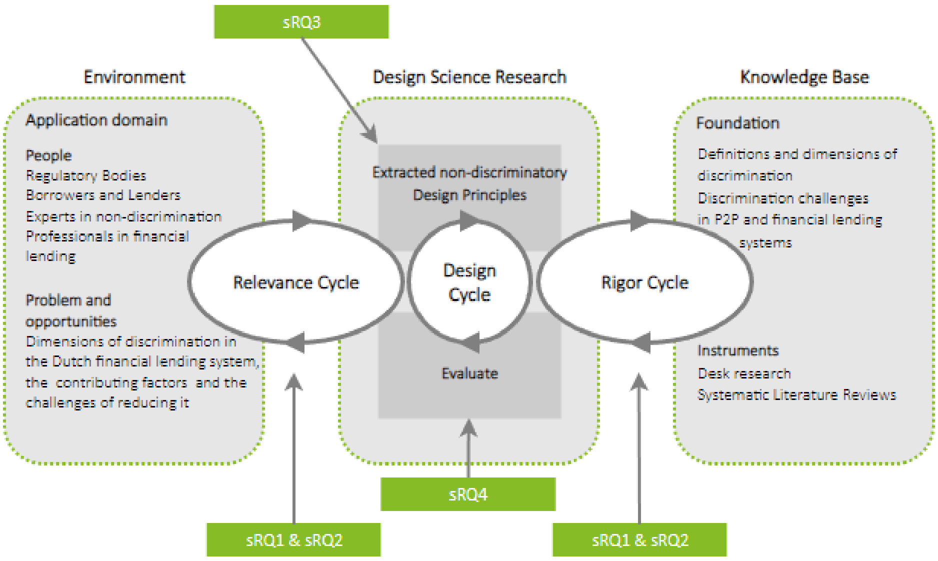


Figure 5: Research approach and sRQ fitted on the DSR cycles by Hevner (2007)

3.1.1 Limitations of DSR

Despite its benefits, DSR also has limitations in generalizability of the results and complexity of the research design. The solutions developed through DSR are often specific to the perception of the participants about the problem which limits the generalizability of the results (Goecks et al., 2021; M. Wang et al., 2011). To address this limitation, this study incorporates multiple perspectives through interviews and considers a wider range of potential solutions. Moreover, the process can also be time-consuming and requires quite some participants for generalizability. To overcome this limitation, the research was scoped down to only focus on the extraction of design principles, rather than the entire architecture. In addition, it narrowed its focus solely to financial discrimination, rather than encompassing the broader concept of financial inclusion. This simplifies the research process and enables further research to use these principles to design the actual architecture.

3.1.2 Sub-questions

This section defines the set of sRQ to answer the mRQ. To structure these sRQ and to provide a logical presentation of the research on deriving design principles, the 'Research Method for Extracting Principles' by Bharosa & Janssen (2015) is used (Figure 6). This method is suitable for addressing complex and ill-structured problems, which often have both social and technical elements, making it a suitable method for the complex multi-actor problem of lending systems. This section aims to provide a high-level overview of the research approach and data collection methods used to answer each sRQ. A more detailed description will be provided in each respective chapter.

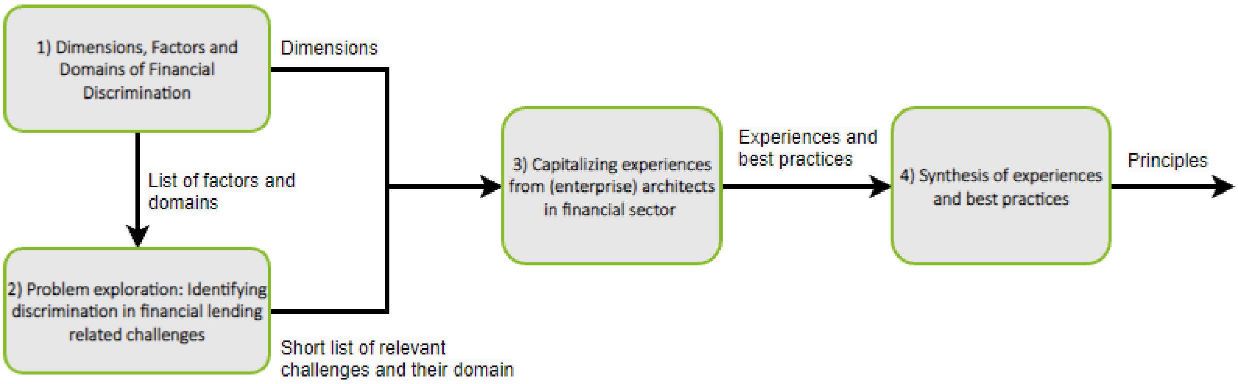


Figure 6: Research Method for Extracting Principles (adjusted from Bharosa & Janssen (2015))

---

1. What are factors contributing to financial discrimination in the financial lending system?

This sRQ explores the factors contributing to financial discrimination in the lending context. As previously mentioned, the direction of this study was revised to encompass a broader examination of discrimination in financial lending, rather than solely focusing on P2P lending. This sRQ can be divided into two parts. The first part focuses on the understanding of what are the dimensions of financial discrimination. Currently, discrimination is defined in the theoretical background section in a quite a generic and abstract level and which is derived from literature. The identification of the dimensions ensured that the concept of discrimination is more applied to the context of financial lending. The second part examines the factors that may contribute to the identified dimensions of discrimination. Analysing these factors can provide a better understanding of the potential domains where discrimination can occur and how they interact. The findings form a basis for this study and ensure that the challenges and extracted design principles are aligned to the dimensions and factors of financial discrimination. In addition, these findings are useful to inform participants during interviews and workshops about the concept of discrimination within the context of this study. This clarity and alignment with the scope of the study ensures that the sessions with participants remain focused and relevant to the research objectives.

In order to address this sRQ, desk research is conducted using existing studies (secondary data) to identify the various dimensions and factors of financial discrimination. Following this, 13 semi-structured interviews are conducted with stakeholders and a wide range of (scientific) experts and professionals with experience in the financial industry and/or financial discrimination.

2. What are the challenges for realizing a non-discriminatory financial lending system?

The second sRQ focuses on identifying the challenges faced in realizing a non-discriminatory financial lending system in the Netherlands. By identifying these challenges, this study aims to provide a deeper understanding of the problem and the underlying reasons behind them. Moreover, the identified challenges will serve as a basis for developing design principles that address these challenges and safeguard non-discrimination in financial lending. This approach aligns with previous research, such as the work of Matheus et al. (2021), who identified obstacles to digital transparency in the literature and linked them to relevant design principles. Similarly, Bharosa & Janssen (2015) explored information quality related problems and interviewed architects on how they tackled these problems in their application design. This study employs a comparable approach to extract the design principles to overcome the challenges.

To answer this sRQ, a systematic literature review (SLR) is performed using keyword searches in the search engine Scopus to identify challenges from existing studies (secondary data). This is followed by semi-structured interviews with the same individuals as sRQ 1 to evaluate whether these challenges are also applicable to the Dutch context and identify additional challenges they have experienced.

3. What are design principles for a non-discriminatory financial lending system?

The objective of this sRQ is to derive the design principles that will guide the development of financial lending systems that ensures non-discrimination. The principles should be generalizable to different situations and provide organizations with guidance on making the right decisions when striving for non-discrimination. The TOGAF template will be used to define these principles, including a name, statement, rationale, and implications for each principle. To extract these design principles, the approach suggested by Bharosa & Janssen (2015) was followed. This approach focuses on capitalizing on experiences to identify the best practices of enterprise architects in designing lending systems with non-discrimination in mind. A total of seven semi-structured interviews were conducted with enterprise architects who have knowledge and experience in the financial sector. These architects were chosen as the most suitable interviewees as they are actively involved in the creation and utilization of design principles in practice. It is important to note that these architects specialize in either the technical or the business aspect of architecture. This distinction is made as it helps to discern possible differences between these two groups regarding their views on how to approach this issue. These architects were mainly contacted through Deloitte's network with architects.

4. To what extent are the created design principles usable for the improvement of non-discrimination in financial lending?

The purpose of sRQ 4 is to evaluate the usefulness of the created design principles. Another group of architects was invited for an evaluation workshop to evaluate the usability of the design principles. Architects are the ones who know the usefulness of these design principles in practice as they are the ones who usually use these principles to design. As such, they are deemed suitable to evaluate the usability of design principles for this purpose. It was decided to include another group of enterprise architects who had not previously participated in the study. This selection was intentional to avoid possible bias towards their own developed principles. In addition, these architects also met the criteria regarding their experience in developing principles and designing architecture for financial organizations. In addition, they have a balanced combination of technical and business backgrounds, unlike the previous group of architects in sRQ 3 who are more specialized in one area.

3.2 Quality criteria for qualitative research

The traditional quality metrics used in quantitative research, such as internal validity, generalizability, reliability, and objectivity, are less suitable for evaluating the quality of qualitative studies (Korstjens & Moser, 2018). Instead, qualitative scholars often use an alternative set of quality criteria to judge whether the findings can be trusted (Lincoln & Guba, 1985). Though many interpretations of trustworthiness exist, the most recognized criteria are credibility, transferability, dependability, and confirmability, as outlined by Lincoln & Guba (1985). Multiple strategies have been recommended by scholars, such as Lincoln & Guba (1985) and Saunders et al. (2009), to address these criteria. This study carefully selects strategies that align with its design to ensure the aforementioned quality criteria are met.

3.2.1 Credibility

Credibility, comparable to validity in quantitative research, addresses the confidence level in the truth of research findings (Lincoln & Guba, 1985). It ensures that the findings are not only plausible but also a correct interpretation of participants' original perspectives. To enhance credibility, this study employed methodological triangulation, collecting data through various methods such as semi-structured interviews and desk research. Additionally, a member-checking procedure was adopted, where participants received a summary of their interviews for feedback, and a subset was consulted mid-study to validate interpretations of the findings. Persistent observation was also included as part of the research methodology. All interviews were transcribed and individually analysed. During data analysis, codes were developed to examine the interview data. Persistent observation further improves the credibility. During data analysis, codes and concepts were established to examine interview data. By repeatedly revisiting and analysing the data, leading to potential modifications in codes and concepts as necessary.

3.2.2 Transferability

Transferability, comparable to generalizability in quantitative research, refers to the extent which qualitative research results can be generalized to other contexts or settings with different respondents (Lincoln & Guba, 1985). Transferability is achieved when readers can relate findings to their own circumstances. To assist readers in making such ‘transferability judgements’, researchers should offer a comprehensive description of the study participants and the research process. This detailed description enables readers to judge the applicability of the findings to their specific settings, as they are best equipped to understand their unique contexts. Accordingly, this research ensured detailed descriptions of the setting, sample, sample size, sampling strategy, expertise description, interview protocol, and inclusion and exclusion criteria, providing the information needed to enable such judgments.

3.2.3 Dependability and confirmability

Dependability, comparable to reliability in quantitative research, concerns the consistency of findings and whether they could be replicated in a similar study. Confirmability, comparable to objectivity in quantitative research, emphasizes a degree of neutrality in the findings, ensuring they are shaped by the respondents rather than researcher biases, motivations, or interests. (Lincoln & Guba, 1985)

A crucial strategy to ensure both dependability and confirmability is the implementation of an audit trail. In this research, a detailed record was maintained, documenting the decisions made throughout the research process, including the stages of data collection, data analysis, and interpretation. By maintaining this record, the research seeks to foster transparency, allowing others to trace its steps and decisions.

3.3 Research Flow Diagram

Each step in the research approach and methods used are visualized in a research flow diagram (Figure 7)

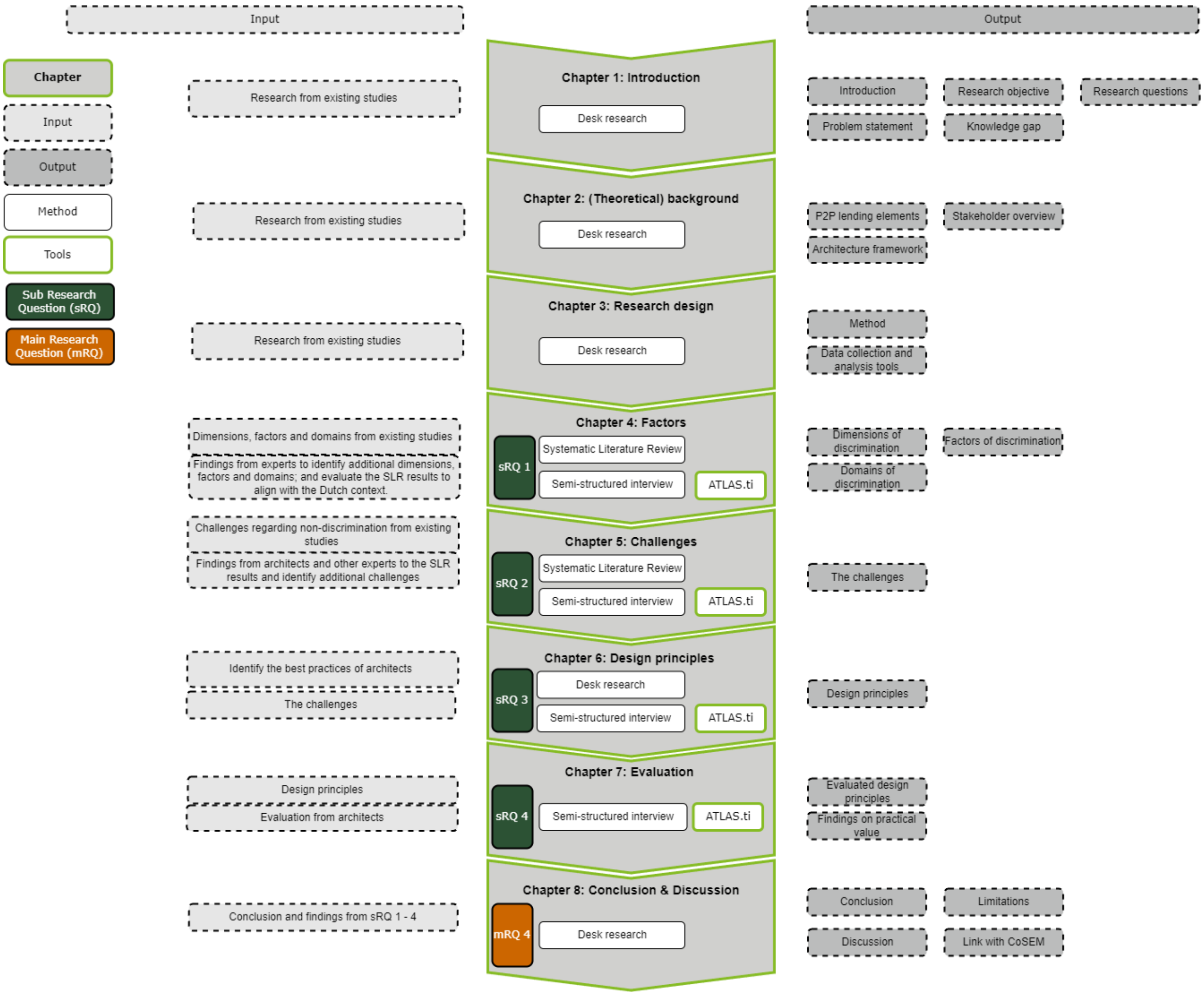


Figure 7: Research flow diagram

4 The dimensions of financial discrimination

So far, the theoretical background chapter provides a generic and abstract definition of discrimination, drawn from existing literature. To increase the applicability of the concept of discrimination in the context of financial lending, it is important to identify and define its specific dimensions. Dimensions in this study refer to the specific ways or perspectives through which the concept of discrimination manifests or can be observed in financial lending. This first step provides a more contextualized understanding of financial discrimination. This chapter then examines the factors that could contribute to the identified dimensions of discrimination. These factors provide a better insight into possible domains where discrimination can manifest and how these factors interact in that context. Thus, factors are the underlying elements or causes that could potentially lead to or exacerbate these dimensions of discrimination, influencing how and where discrimination is manifested in the financial lending system.

The findings of this analysis form the fundamental basis for understanding the following chapters. In addition, these dimensions and factors are useful in informing participants to ensure that the interviews remain focused and relevant to the research objective.

These lead to the following sub question that will be answered:

1. *What are factors contributing to financial discrimination in the financial lending system?*

This study was initially designed to exclusively examine the architecture of P2P lending. However, closer examination revealed that the problem of discrimination should not be limited to this practice. Interviews conducted as part of this chapter revealed that discrimination is prevalent throughout the entire lending system in the Netherlands, indicating that it is a systemic issue rather than solely a P2P lending architectural concern. Consequently, a broader perspective was considered more valuable in understanding the factors of discrimination. It is therefore important to note that whereas the desk research in this study is still primarily focusing on P2P lending, the interviews adopted a wider scope, aiming to gain insights into the Dutch financial lending context and its associated discrimination issues.

4.1 Summarized approach

To answer the first sRQ, the study used a combination of desk research and interviews. Desk research provided an overview and review of existing literature related to the topic which built a strong foundation for developing interview questions and identifying gaps in the literature. Interviews collected primary data from various perspectives on this issue enriching and contextualizing the desk research findings to the Dutch context. Moreover, it can gather new insights to fill the gaps found in the literature. The combination of desk research and interviews allowed triangulation of data, which cross-checked information from multiple sources to ensure the rigor and accuracy of findings. This approach also leveraged of each method's strengths and offset their limitations. For example, desk research provided a broad overview of existing literature, while interviews provided in-depth insights into specific perspectives or experiences in the Dutch context.

4.1.1 Desk research

To answer the first sRQ, desk research is conducted to get an understanding of the following areas:

- Examining the primary factors that contribute to discrimination in P2P lending, based on existing literature. The aim is to:
    - Identify gaps that have not been adequately explored in existing literature.

    - Identify consensus and patterns related to specific discriminatory factors by comparing the existing findings.
    - Prevent duplicating works that have already been conducted in the field.
    - Understand what discrimination is in the P2P lending context.
- Identifying the role, motivations, and concerns of borrowers and lenders using P2P lending platforms. The aim is to:
    - Understand why users choose to use P2P lending platforms.
    - Understand how users' actions could contribute to discrimination.

To explore the perspectives of lenders and borrowers in the context of P2P lending, a literature review was conducted using the Scopus search engine. The search terms used were: (("P2P lend*" OR "peer-to-peer lend*") AND "lender*" OR "borrower*" AND ("motivation" OR "concern*" OR "role*" OR "intention*" OR "behavior*" OR "statistical" OR "taste-based" OR "discrimination"))

The search was conducted using title, abstract, and keywords to narrow down the number of articles. This initial search yielded 224 articles published up to March 25, 2023. The article set was then filtered based on inclusion and exclusion criteria, resulting in a final selection of 33 articles. All 33 eligible publications underwent a thorough manual review to assess their relevance to the research topic. After evaluating the full texts of these articles, the number was further reduced to 8 articles that met the specific requirements and objectives of the study. The list of selected articles, are provided in Appendix B. The studies by Yang & Lee (2016) and Prystav (2016) were identified as the most relevant in exploring the discrimination factors in P2P lending. These articles served as the starting point for the research, and a forward and backward search was conducted to identify additional studies. This additional search resulted in the inclusion of 9 more papers, bringing the total number of papers included in the review to 17.

4.1.2 Interviews

In this comprehensive study, a total of 13 participants were carefully selected to ensure a diverse range of perspectives and expertise. The details of the participants can be found in Table 2. The study initially conducted 16 interviews as part of the research process. However, it was discovered during these interviews that three potential participants who were initially perceived as experts did not possess the necessary expertise and experience to provide valuable insights. As a result, these individuals were excluded from the study, and their perspectives were not included in the analysis and findings. This exclusion ensured that the analysis focused on relevant and knowledgeable perspectives from participants who could contribute meaningful insights to the research. The process of determining someone as an expert is described in Appendix B.

The primary objective of the interviews was to gather a comprehensive understanding of the discriminatory factors involved in lending practices. To achieve this, a diverse group of individuals representing various key stakeholders was interviewed. The interview protocol can be found in Appendix C. To capture the experiences and perspectives of borrowers and lenders, professionals from the Netherlands Institute for Human Rights, a Dutch organization dedicated to combating discrimination, as well as the chairman of a foundation supporting alternative financing solutions for entrepreneurs, were interviewed. These individuals have extensive interactions with Dutch citizens involved in borrowing and lending, which provided valuable insights into the perspectives and issues faced by these stakeholders. Furthermore, professionals representing different layers of the architectural framework outlined in the Theoretical Background chapter were selected to represent financial organizations. For example, an expert with specialized knowledge and experience in credit scoring models and algorithms used in financial lending was interviewed. This approach ensured a thorough examination of the potential factors in each layer of financial organizations. Finally, to gain insights into the perspectives of regulatory bodies, interviews were mainly conducted with lawyers, jurists, and researchers. These individuals had a deep understanding of the legal context surrounding financial discrimination, particularly within the Netherlands. They also provided information on existing policies, ongoing efforts to combat discrimination in the financial lending sector, and offered insights into recent developments, emerging trends, and potential gaps in current understanding.

It is important to note that while this study acknowledges that not all possible factors of discrimination have been collected, the selected perspectives provide a socio-technical view that enables an understanding of the various factors contributing to discrimination.

Table 2: Overview of the interviewees and their expertise

ID	Date	Expertise
A1	06-04-2023	Member of the UN Committee on the Elimination of Racial Discrimination and researcher in human rights
A2	06-04-2023	Investor with 7 years of experience in P2P and pool lending
A3	12-04-2023	Researcher specializing in fairness, bias, and discrimination in machine learning algorithms
A4	13-04-2023	Director of a European crowdfunding organization and the chairman of a foundation dedicated to supporting alternative financing solutions for entrepreneurs
A5	14-04-2023	Specialized jurist with expertise in labour and discrimination law, and a member of the EU network of Legal Experts
A6	17-04-2023	Legal professional at a Dutch organization dedicated to combating discrimination
A7	18-04-2023	Experienced professional in UI/UX design for financial organizations
A8	18-04-2023	Senior data architect in the financial sector
A9	18-04-2023	Experienced in commercial banking, payments, and fintech across the Netherlands and Europe
A10	19-04-2023	Senior enterprise architect in the financial sector
A11	21-04-2023	Experienced in developing and implementing solutions in the lending industry
A12	24-04-2023	Researcher and practitioner in trustworthy AI and credit scoring for financial organizations
A13	01-05-2023	Member of the Netherlands Institute for Human Rights with a focus on the intersection of digitalization and human rights

4.2 Main findings of Literature

The existing literature on discrimination in P2P lending primarily examines lenders who discriminate based on information provided by the platform and borrowers themselves. To gain a better understanding of what is currently researched on discrimination in P2P lending and the variables that lenders discriminate borrowers on and their possible impact, a literature review was conducted. Appendix E shows the table with the literature review and the information on the objective of the paper regarding discrimination in P2P lending, the type of information, the tested variables by the researcher and the impact discovered. However, it should be noted that most of the literature focuses on P2P lending in China, the UK, and the US, with limited research available on P2P lending in the Netherlands and the broader EU context. The lack of studies specifically on P2P lending in the Netherlands was not addressed in the literature. The interviews later provided valuable insights that shed light on this knowledge gap related to discrimination in P2P lending in the Netherlands. These insights will be discussed in detail in this chapter.

4.3 Main findings of Interviews

This section summarizes the main findings derived from interviews. The summary of each interview can be found in Appendix D. In the synthesis section, the information gathered from multiple sources, including literature and interviews, will be synthesized to explore the relationships between the various views and arguments presented.

4.3.1 Atlas.ti Analysis

This study employed an abductive reasoning approach, a method that seeks the most plausible explanation for a set of observed phenomena. This approach is especially suitable for complex situations where existing theories fall short of providing clear explanations. Given the study's novelty and the lack of a pre-existing framework, abductive coding facilitated the identification of categories within the data and the subsequent development of codes derived from these categories. Moreover, the study adopted an architectural framework for participant selection, streamlining and structuring the selection process

Atlas.ti was employed to analyse the data gathered from the 13 interviews conducted. The interviews were transcribed and subjected to a three-stage coding process: open coding, axial coding, and selective coding. In the open coding stage, initial codes were generated to capture the key concepts and ideas (quotations) present in the data. The axial coding stage focused on establishing connections and relationships between the codes to organize the initial codes into broader categories. The selective coding stage focused on the refinement and selection of the most relevant codes that directly addressed the sRQ.

In total, there were 317 quotations that were analysed and categorized into 108 codes. Out of these codes, 20 represented factors, 5 represented stakeholders, 4 represented dimensions, and 6 represented domains. In this section, 37 codes were examined and discussed. The remaining codes were related to the challenges and solutions provided by the experts, which will be addressed in the next chapters of the study.

4.3.2 Overview of codes

The network diagram presented in Figure 8 shows the interconnections among the codes utilized in the coding process of Atlas.ti. These codes and relationships are primarily identified by the interviewees during the interview process. For instance, one interviewee pointed out that utilizing complex risk models can result in more precise calculation of risk and credit scores. However, the complexity of such models can also result in decreased transparency due to the backbox, making it difficult to explain how the score is determined. This demonstrates a contradictory relationship between accuracy and transparency.

The network includes 37 codes that cover five main stakeholders: regulatory bodies, financial organisations, P2P lending, borrowers, and lenders. These stakeholders are presented in green. Additionally, 25 subdimensions were identified in the study, which are presented in light red. These subdimensions were further categorized into six main categories, presented in blue, which include regulations, human, and four layers of the architecture framework (user interface (UI), business, data, and technology) that can lead to discrimination, presented in dark red. The application layer was not explored further in this study because limited or no discrimination dimensions were found in both the literature and the interviews.

Figure 8 aims to illustrate the complexity of the problem, based on the various perspectives and insights gathered during the interviews. It is important to note that this figure does not aim to provide clarity to readers but rather serves as a visual representation of the complexity and interconnectedness of the issue.

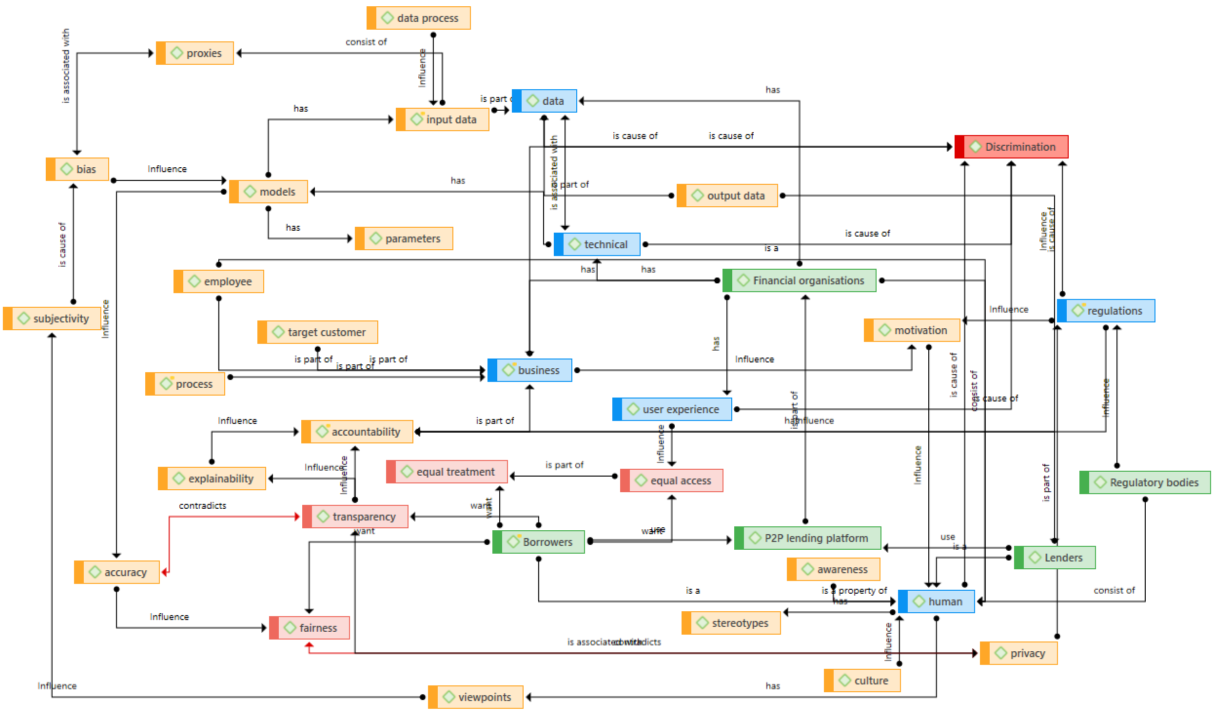


Figure 8: The dimensions of financial discrimination in the Dutch P2P lending context. Green: stakeholders, blue: domains, orange: factors, dark red: discrimination and light red: dimensions of discrimination

A diverse range of perspectives regarding discrimination have been gathered, revealing numerous dimensions that could contribute to discriminatory practices. These dimensions are interrelated as shown in Figure 8, such that changing one aspect can have impacts on other dimensions and are often shaped by the various stakeholders involved. The figure will be divided and elaborated on in more details in the next section.

The overview is useful as many interviewees emphasized the importance of a comprehensive understanding of the various dimensions contributing to discrimination. They noted that researchers and stakeholders tend to focus on one domain while overlooking others, leading to an incomplete overview of the issue. Interviewee A13 stressed the importance of considering all dimensions that can lead to discrimination, rather than just focusing on the legal or technical aspects. Moreover, Interviewees A1 and A5 highlighted that combatting discrimination is not only about complying with laws and regulations but also addressing dimensions such as stereotypes and awareness. A6 pointed out that organizations are often more reactive than proactive in preventing discrimination, acting only after discrimination has occurred. This is possibly because they lack knowledge on how to prevent it. This opinion was shared by several interviewees (A4, A7, A8, A10, A11 and A13). These findings indicate the relevance of this study in practice, as it considers various perspectives to gain a broader understanding of the different dimensions of discrimination and their relationships. Furthermore, this study also aims to develops principles to prevent discrimination.

4.3.3 Simplified framework

To enhance the simplicity and understandability of the mapped complexity, a framework is constructed, as shown in Figure 9. The arrows in the framework indicate how each factor can influence one another and the dimensions of discrimination.

To make the concept of discrimination more operational and applicable in the context of (P2P) lending, it will be divided into four dimensions. These dimensions will help provide a clearer understanding of discrimination, as it can vary for each person and context. By using dimensions, there is also a better scope for the study, as it can focus on the main dimensions rather than attempting to cover every aspect of discrimination. The use of dimensions also aids in identifying relevant factors that impact each dimension. To clarify, if a factor has an influence on one of the dimensions, it can be considered as a relevant factor to be taken into consideration. During the desk research and interviews, four primary dimensions of financial discrimination in lending are commonly discussed: differential treatment, differential outcome, unequal access, and lack of transparency.

- Differential treatment refers to intentional discrimination based on protected characteristics.
- Differential outcome occurs when there is no explicit intent to discriminate, but the treatment results in unequal lending outcomes.
- Unequal access refers to exclusion from loans due to platform design, occurring even before the loan process starts.
- Lack of transparency refers to financial organizations not providing explanations for loan outcomes, making it challenging to prove discrimination.

These dimensions capture the key areas where discrimination often occurs in the lending context. The identification of these dimensions allows for a more comprehensive exploration of relevant factors that influence each dimension. Thus, any factor that affects one of these dimensions is considered relevant in the study. It is important to note that while the dimensions of discrimination are discussed, this study does not provide an in-depth examination of each dimension. Rather, the focus is on identifying and explaining the factors that contribute to discrimination within these dimensions which are presented in more details in section 4.4.1.

For clarity, the arrows do not directly connect to the dimensions of discrimination but are represented by different colours.

- Green indicates factors that can influence all three dimensions of discrimination.
- Orange mainly impacts the dimension of unequal access
- Blue mainly affects the dimension of unequal treatment
- Purple mainly relates to unequal treatment
- Yellow mainly relates to the lack of transparency dimension

It is important to note that there are multiple arrows in the framework, representing the connections between factors based on insights from the interviews. However, this study does not provide an in-depth analysis of the exact influence and significance of each factor on the others. Further research could delve into each arrow and explore the relationships between factors more extensively. Unfortunately, due to time constraints, this level of analysis was not possible in the current study. The limitations of this study will be discussed further in the limitation section.

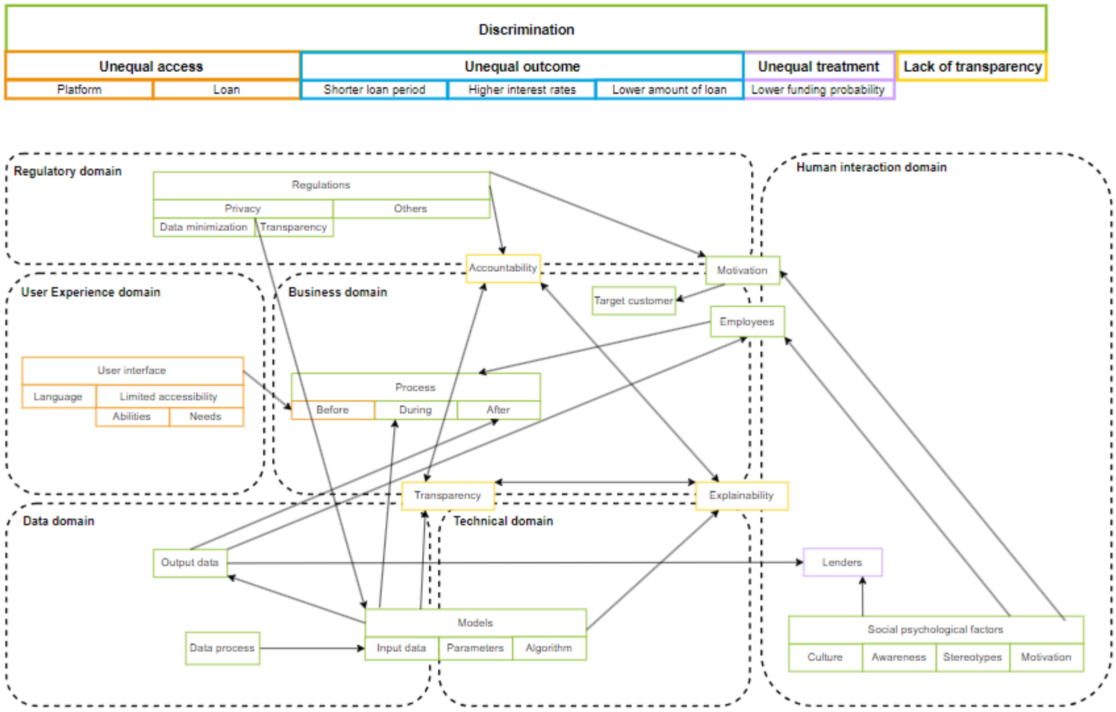


Figure 9: Overview framework consisting of the dimensions, domains and factors of discrimination

To understand the figure and potential meanings of the arrows, let's consider the following situation which explains a series of examples. A more detailed description of each domain and its factors will be explained in the upcoming sections. Regulations play an important role in influencing the motivation and accountability of organizations. When there are regulations that enforce non-discrimination guidelines, such as fines for non-compliance, organizations are more incentivized and motivated to prioritize non-discrimination efforts. However, currently, the focus of regulations in the financial sector is often on other areas such as privacy and sustainability, which may overshadow the importance of non-discrimination. The motivation of organizations also affects their target customer focus. If certain groups of people are not perceived as the most profitable customers, organizations may not invest in designing UI and processes that meet their specific needs and abilities. This can result in unequal access to financial services, particularly for vulnerable groups such as the elderly, disabled individuals, or non-Dutch speakers, who may face barriers in accessing loans due to the design of the platform.

Another example is that during the business process, financial lending organizations, including P2P lending platforms, rely on risk models to calculate loan amounts and interest rates. These models involve input data, parameters, and algorithms that interact to generate an output used to determine a borrower's creditworthiness. The lack of transparency in these models makes it challenging to trace how certain decisions are made. If complex algorithms are used that are accurate but not easily understood, it can impact the explainability and accountability of the models. Organizations may struggle to explain the outcomes of the model and may evade responsibility for the results. Additionally, social-psychological factors contribute to the problem. Employees involved in the design process may unknowingly incorporate biases throughout the entire process, which can result in unintentional discrimination. Lenders, who have access to the model's output, such as interest rates and loan amounts, may also be influenced by their own biases when making lending decisions. The decision-making process of lenders is currently a black box, as they are not required to explain why they approve certain loan requests and not others. This means that exact reasons behind the decision-making process are not identifiable, leading to potential disparities in loan approval rates and lack of transparency.

These factors collectively contribute to the challenges and potential discriminatory outcomes in the financial lending process. The next section will explain these factors in more details along with their corresponding source.

4.4 Synthesis

In this section, the outcomes derived from both the literature review and semi-structured interviews are discussed. Various studies and interviewees offered diverse perspectives on the potential dimensions and factors contributing to financial discrimination.

4.4.1 The dimensions of discrimination

The concept of discrimination is broad and abstract, and the current definition used in this study is primarily derived from literature and the way it is described in Dutch law. The interviews (A1, A4, A6) also explained that the practical application of the anti-discrimination law is not easy, and that subjective interpretation can vary depending on the point of view of the observer. A6 and A13 both discussed that indirect discrimination complicates the matter because it is not always immediately visible on the surface, making indirect discrimination difficult to prove. Moreover, indirect discrimination is often unintentional, meaning that the responsible person is not aware of his discriminatory acts.

4.4.1.1 Unequal outcome and treatment

During the interviews, A3 provided an explanation of discrimination, highlighting two dimensions: differential treatment and differential impact. Differential treatment refers to intentionally treating individuals differently based on protected characteristics such as race. This means that person A and person B are intentionally treated differently due to their race, for example. On the other hand, differential impact refers to situations where there may not be an explicit intention to treat people differently based on characteristics beyond borrowers' control. However, the outcome or result of the treatment leads to differences in lending outcomes. For example, using education level as a criterion for approval can unintentionally discriminate against those with less access to quality education. Even if the differential impact is not justified by a legitimate business reason, it is still considered discrimination. A3 also noted that determining what qualifies as a legitimate business justification can be subjective. This observation is supported by A5, A6 and A13.

Furthermore, it was observed that the existing literature on discrimination in P2P lending, as shown in Appendix E, predominantly focuses on the unequal treatment of borrowers by lenders. Lenders often discriminate against borrowers based on protected characteristics, driven by their own biases and perceptions of what they deem trustworthy. Studies such as Pope & Sydnor (2011) have revealed that factors like race or age can lead to inequalities in loan success probabilities. A4 provided an explanation, suggesting that these inequalities can be attributed to biases and stereotypes held by investors, influencing their decisions, and resulting in unequal outcomes for certain groups. This implies that unequal treatment often leads to unequal outcomes.

In addition to unequal treatment, several studies have highlighted inequalities in loan outcomes, including higher interest rates, lower loan amounts, or shorter loan periods for certain individuals compared to others. These inequalities are often determined by the risk models employed by financial organizations. A10 mentioned that numerous business rules are utilized to calculate an individual's risk, making it challenging to identify the specific factors that contribute to particular loan outcomes. Moreover, A3 mentioned that if there is a difference in loan outcomes between men and women, it can be explained by the income disparity caused by women working part-time. They acknowledge societal debates about whether this is a choice or a structural issue. From the bank's perspective, giving a loan to someone who cannot afford it would be financially harmful. They argue that this difference in outcome is not unfair discrimination, but rather a result of income disparities. They also mention the complexity of cases where indicators of loan outcomes and risks have a correlation with legally protected characteristics.

These findings align with the theoretical background chapter, where the legal perspective on discrimination distinguishes between direct and indirect discrimination. Unequal treatment corresponds to direct discrimination, while unequal outcomes can be seen as indirect or systematic discrimination, as discussed by Daniel et al. (2014).

4.4.1.2 Unequal access

During the interviews, it was clear that many interviewees recognized unequal treatment as a form of discrimination. However, there was also a strong emphasis on unequal access to loans as another dimension of discrimination. A7, for instance, argued that the platform itself may not be designed to accommodates certain individuals' needs and abilities, resulting in their exclusion from accessing the platform and loans. In contrast to governments, which have a mandatory duty and motivation to ensure equal access for all, financial organizations may not have the same incentives as it would cost them money. This point was further supported by A6 and A13.

Additionally, A4 pointed out that the specific type of product or service provided by financial organizations can lead to exclusion when it does not align with the needs of certain groups. C4 provided an example, highlighting that most loans have a fixed duration and include an interest rate component in the Netherlands. However, there is a significant population of approximately one million Muslims in the Netherlands who prefer to finance Halal, which means without the inclusion of any interest payments. It was expressed as surprising that no financial organisations in the Netherlands currently offers products that cater to this preference, given the size of the target group compared to the size of the Dutch population. This highlights a form of exclusion based on cultural or religious practices. While this aspect of exclusion is relevant, it will not fall within the scope of this study. This is because the interest rate is a key component of (P2P) lending as discussed in the Theoretical Background Chapter.

The observation that dimensions of discrimination can occur even before the loan process begins, rather than solely during or at the end of the process, highlights an important aspect that may not be widely discussed in the P2P lending literature.

4.4.1.3 Lack of transparency

The literature on P2P lending emphasizes the importance of transparency in addressing information asymmetry between lenders and borrowers. Transparency in P2P lending context means that the more information lenders have about borrowers, the less information asymmetry exists. The transparency regarding information asymmetry will not be used as a dimension of discrimination in this study as increased transparency (i.e., providing more borrower's information to lenders) can potentially enable lenders to discriminate based on additional characteristics. This aspect will be further discussed in this chapter.

The lack of transparency referred to in this study is primarily related to the openness and clarity of the lending process itself. Financial organizations should provide clarity on how they arrive at specific loan outcomes. Moreover, they should be able to explain to other stakeholders how these outcomes are derived. This level of transparency is important because many interviewees in the study (A1, A5, A6, A7, A13) highlighted that one of the challenges with discrimination is the difficulty of providing evidence due to the lack of transparency. Borrowers often struggle to prove that they have been discriminated.

This shows that transparency can have different levels and nuances. However, due to time constraints and the primary focus of this study, the exploration of these nuances and levels of transparency was beyond its scope. In this study, the lack of transparency refers specifically to the financial organization's lack of transparency about how they arrive at certain outcomes and the lack of explain ability towards borrowers as well as other stakeholders (e.g., regulatory bodies). This lack results in a situation where the financial organization avoids taking accountability by not being able to explain how certain decisions are made, as noted by A6, A11, and A12.

4.4.2 Dutch financial system

A4 and A9 pointed out that the system of financial loans differs from country to country due to differences in stakeholders and legislation. Therefore, this section will provide an overview of the Dutch financial system to understand the Dutch context before examining the different factors of discrimination.

- The Netherlands has no credit bureaus like other countries, only Bureau Krediet Registratie (BKR) which is a monopoly on the Dutch market. BKR maintains the Central Credit Information System (CKI), which tracks all credit information for individuals and businesses throughout the Netherlands.
- There are also no credit scores as there are no credit bureaus in the Netherlands, so Dutch financial organizations need to develop their own credit risk model.
- The Netherlands has strict rules regarding interest rates, so lenders cannot charge too high interest on loans.
- There are a limited number of P2P lending platforms active in the Netherlands.
- Compared to other countries, obtaining a loan for mortgages in the Netherlands is relatively easy
- There are fewer opportunities to get a loan for personal use.

4.4.3 Borrowers

This section focuses on borrowers who are the individuals primarily affected by discrimination in the (P2P) lending context, rather than being the factor of discrimination. Understanding their perspective is relevant as dimensions like unequal treatment and unequal outcomes largely impact them. Consequently, it becomes important to understand their motivations for using P2P lending platforms and address their concerns. This ensures a better understanding of the challenges they face.

4.4.3.1 The motivation of borrowers

When a new service can provide greater value to a customer compared to existing services, through improvements in economic benefits, convenience, and satisfaction, it is perceived to be beneficial (Yang & Lee, 2016). According to a study by Yang & Lee (2016), online P2P lending is one such service that offers advantages over traditional financial institutions, such as relatively high-interest rates and greater convenience. However, traditional financial institutions also have their own advantages such as relying on personal interactions to avoid impersonality and anonymity, as well as offering better security through offline channels. Despite this, if people perceive that the online P2P lending service offers more value than other options, they may develop a positive attitude towards the service even without prior experience (Yang & Lee, 2016). P2P lending can provide several benefits to borrowers, such as lower interest rates, lower transaction costs, faster loan approvals, and flexible loan terms (Klafft, 2008; Lenz, 2016; Yang & Lee, 2016). Borrowers who do not have a well-verified credit score can often provide alternative information and still get a loan (Ge et al., 2017). Borrowers can also typically cancel loan contracts prematurely without paying a prepayment penalty (Lenz, 2016). P2P lending offers lower transaction costs due to 24/7 accessibility of the platform, reduced bureaucracy and documentation requirements, and a simple and transparent application process.

4.4.3.2 The concerns of borrowers

There are also concerns for borrower such as the uncertainty over whether P2P lenders will accept their loan offer and whether they will be able to borrow the full amount requested (Lenz, 2016). P2P lenders do not provide an upfront guarantee that they will finance a loan at a certain volume, interest rate, credit risk, and maturity. If lenders refuse to finance a loan request, the offer is often improved by increasing the interest rate until it meets the lenders' requirements (Lenz, 2016). Lenders prioritize reducing their risks and maximizing their return on investments, which influences their decision to provide loans only to loan requests that meet their, often subjective, requirements. This will be discussed in more details in the Human Interaction domain section. This decision-making process is not solely based on hard information but also on soft information which leads to higher interest rates, lower loan amounts, or even denial of loan requests for disadvantaged borrowers.

Borrowers could perceive this as unequal treatment, as lenders could base their decision-making on subjective factors that may not reflect their true creditworthiness. A2 made an interesting point that borrowers cannot verify whether lenders are engaging in discriminatory practices since lenders are not required to disclose their decision-making process to borrowers. This creates a lack of transparency on the part of the lender, making it difficult for borrowers to understand why lenders are not funding their loan applications and whether this is based on information and characteristics beyond their control. As a result, borrowers may feel they are not being treated equally. A2 stated the following:

> *"This is tricky because borrowers cannot check whether a lender is discriminating, this is happening outside the platform and inside someone's head. Plus, lenders do not have to explain why they do not finance a particular loan."*

A6 and A13 stated that what is considered discrimination differ for each person due to their culture and viewpoints. According to A2 and A4, lenders and investors do not perceive themselves as discriminatory, but rather view their lending decisions to minimize risk and protect their own interests. A2 also emphasized this point, stating that:

> *"I am sure that most [lenders] do not consciously think that they are being discriminatory by financing to certain projects, we simply want to protect ourselves and run less risks and get a higher return, that is more of our priority by joining these kind of platforms."*

The study by Lenz (2016) also showed that P2P lenders are motivated to invest only when the financial returns from using platforms compensate for the additional risk taken and surpass other investment opportunities of similar risk. Lenders have been observed to primarily concentrate on their return on investment and risk reduction, often overlooking potential discriminatory practices. Additionally, the perception of discrimination can be subjective. While some individuals view it as discriminatory, others may consider it a reasonable measure to mitigate risks. This shows that perceptions of discrimination can vary greatly depending on one's position and perspective. It underlines the need for objective criteria to define what constitutes discrimination in the context of lending.

In addition to concerns about potential discrimination from lenders, borrowers have also expressed criticism regarding the lack of transparency in how lending platforms assess creditworthiness. Lenz (2016) highlighted that borrowers may not have visibility into the types of data used by platforms or how credit ratings are calculated. This lack of transparency raises concerns about potential discriminatory practices. A6 and A13 emphasized the importance of transparency to borrowers. However, they noted that transparency is only valuable if borrowers can understand the information provided. Even if an organization claims to be transparent, if the borrowers cannot comprehend the information, it is still considered as not transparent to them. This shows the distinction between transparency and explainability. Organizations often choose not to disclose their algorithms for competitive reasons, as these algorithms can serve as important differentiators from their competitors. Additionally, the complexity of these algorithms can make it challenging to communicate their workings in a manner that is easily understandable to the public. In some cases, even the companies themselves may not have a full understanding of the inner workings of their algorithms (A12).

4.4.4 Regulatory domain

This section provides an overview of the various factors and influences within the regulatory domain.

4.4.4.1 Influence of regulations on lending market in the Netherlands

A2, an experienced Dutch investor in P2P lending platforms, expressed interest in using Dutch P2P platforms but noted that there are not many available. A4 provided the following rationale for this:

> *"I think the most important reason why P2P lending is not big in the Netherlands is simply the restriction in the legislation, or rather the consumer protection in the legislation, which requires interest rates to remain relatively low."*

A4 highlighted that consumer lending in the Netherlands is subject to strict regulations, specifically related to the maximum interest rate that can be charged. Consequently, many P2P lending companies have found it unprofitable to operate in the country. In contrast, countries like the United States do not have a maximum interest rate for lending. A10 explained why the Netherlands has such regulations. Historically, the country had "flits kredieten" which were short-term loans that could be obtained at high interest rates. However, the government found that this was not good for consumers' financial well-being, which led to the development of stricter regulations regarding lending.

In addition, both A3 and A4 have highlighted that regulations are constantly changing. A4 noted that the current EU rules on crowdfunding mainly apply to financing for businesses, but there are plans to extend it to consumers in the future. Meanwhile, A3 discussed the EU AI Act, a new set of regulations that will govern high-risk AI systems. The draft of the AI law includes requirements such as credit risk assessments and independent assessments of fairness and transparency for these systems. These examples demonstrate that regulations can and will continue to evolve over time to keep up with industry advancements and emerging challenges. It is also noted that regulations often lag technological advancements and emerging issues. This lag can enhance or create barriers for innovations like P2P lending to flourish in certain countries, such as the Netherlands.

4.4.4.2 The influence of regulations on motivation and priorities of organizations

A11 emphasized the impact of regulation on the task prioritization of financial organizations. Since organizations are often reactive rather than proactive, they prioritize tasks based on regulatory demands. This is because financial organizations have a limited capacity and have to deal with numerous compliance and regulation-driven changes within a specified timeframe. For instance, current strict regulations on privacy, KYC, and ESG have caused financial organisations to prioritize these issues over others. As a result, issues such as reducing bias in risk models have become less of a priority as more resources are allocated to urgent regulatory compliance. A7 further clarified this point by explaining that when organizations face regulatory burdens, such as receiving fines, they become more focused and motivated to take accountability or prioritize the task at hand. This is also noted by several other interviewees (A1, A3, A7, A8, A11).
A8 and A5 highlighted the role of financial considerations in how organizations prioritize improving accessibility for certain groups. A8 pointed out that organizations tend to prioritize customers who generate the most revenue (target group). A7 further emphasized this perspective and stated that:

> *"Only when certain groups move into more mature fragments of the market, it will start to become important to get things done. That is why there are different parts of the market that are treated differently."*

However, this can result in neglecting the needs of underrepresented groups that may not bring in as much revenue. The focus on profitability may overshadow the importance of ensuring equal access and opportunities for all individuals, regardless of their financial potential. Regulatory bodies also face difficulties in enforcing strict control over all issues due to their limited capacity. A11 highlighted this issue by stating:

> *"The regulators face the same problem. They also have limited capacity, so they focus on things that affect the largest population or pose the greatest risk to the financial system."*

This is also expressed by A1 which noted that discrimination is a complex issue that has not received equal attention as it does not affect everyone. However, A1 went on to argue that discrimination should concern everyone since it has an impact on society as a whole.

4.4.4.3 Privacy compared to discrimination

Privacy was often mentioned and compared with discrimination during the interviews. A1 observed that discrimination does not receive as much attention as privacy issues do. This observation was shared by other interviewees, including A8 and A11, who noted that this is particularly true in the financial sector. They highlighted that discrimination is a less concrete issue compared to privacy. For instance, A11 mentioned that there are many practical and legal guidelines available regarding privacy, such as GDPR, while there is a lack of such guidelines for discrimination. This lack of guidance on discrimination was also mentioned by other interviewees (A2, A7, A8, A9, A11, A12). The absence of concrete regulations and guidelines for addressing discrimination can be seen as a causal factor that leads to the limited actions taken to address discrimination in financial organizations. This highlights the relevance of this study in providing more concrete principles to tackle discrimination in the context of financial lending.

A6 pointed out that there is a difference in the legal weight behind the two issues. While discrimination is illegal in many jurisdictions, the decisions of organizations such as the Netherlands Institute for Human Rights are not legally binding. On the other hand, for the value of privacy, GDPR is a legally binding regulation that sets rules for the usage of personal data. Companies that handle personal data should comply with the GDPR regulations and failure to comply can lead to legal consequences. This means that discrimination may be illegal, but it lacks the concrete legal enforcement that the GDPR does have. This can make it more challenging to address discrimination compared to privacy issues. This could also explain why companies prioritize privacy issues over discrimination.

4.4.4.4 The influence of privacy on other factors

Organizations are ensuring that their lending practices follow the privacy rules such as GDPR due to the legal consequences. However, by strictly following these privacy regulations can also lead to other issues related to discrimination. A12 raised an important issue regarding the impact of privacy on various factors of lending, including transparency, accuracy, and fairness. While privacy is essential for maintaining personal safety and dignity, it can also create challenges for fair lending practices. For example, GDPR regulations require organizations to protect borrowers' personal data, which can make it difficult to detect and address indirect biases present in risk models. In addition, accuracy is important to ensure that lending decisions are fair, as it allows organizations to accurately and correctly calculate an individual's creditworthiness. However, accuracy can also conflict with transparency, especially when using complex algorithms to improve accuracy. To clarify, A12 gave the following example:

> *"You can use a complex deep neural network for more accuracy, but then the decision-making process or the mathematics are very complicated, so you cannot be very transparent about it because you actually do not know what happened. You can estimate it, but you cannot explain the workings of a deep complex model to many people that are not highly educated in that area."*

This can lead to situations where customers and organizations do not understand how a loan decision was made, which goes against the transparency requirements of the GDPR for the processing of personal data. The lack of transparency in loan decision-making can create problems with understanding how decisions are made and ensuring accountability. This challenge will be discussed in more details in the next chapter.

4.4.5 Human interaction domain

This section provides an overview of the various factors and influences within the human interaction domain.

4.4.5.1 Social psychological factors

4.4.5.1.1 Stereotypes and culture

According to A1, combating discrimination involves more than just legal measures and stated that:

> *"Combating discrimination is never an issue alone which has to do alone with law, it's just one of these aspects. I think that combating discrimination requires combating against stereotypes and stereotypes will always be there."*

In addition, A4 also noted that the male-dominated industries in the Netherlands often lead to the emergence of stereotypes and biases that favour masculine traits and qualities in defining entrepreneurial success. This can disadvantage women and other marginalized groups. The study by Harkness (2016) also recognized this issue in P2P lending and stated that the differential treatment of lenders may be attributed to stereotypes and biases. A6 and A13 both explained how cultural background and individual viewpoint can lead to these biases and stereotypes. For example, if a person grows up in an environment where certain groups of people are consistently portrayed negatively, they could develop a biased perspective that affects how they view those individuals. Therefore, views on certain issues are subjective and can lead to biases.

4.4.5.1.2 Lack of awareness

A4 shared their observation that many investors tend to believe that they give everyone equal opportunities to invest in. However, a closer examination of their investment practices often reveals the opposite. They stated the following:

> *"When I am in a room with investors and this topic comes up, they all say everyone gets an equal opportunity and they invest in everyone. But if you just look at the numbers, you see something completely different."*

This implies that investors are not fully aware of their own biases against certain groups. Similarly, A12 noted that it can be challenging to become aware of your own biases. This is especially true when individuals are focused on completing other tasks or have competing priorities, such as in a work environment. In such contexts, biases can operate unconsciously and go unnoticed. Other interviewees, such as A10 and A11, also acknowledged that biases are often unconscious.

4.4.5.2 Lenders

This section focuses on lenders as a factor of discrimination within the human interaction domain. To understand lenders, it is important to first examine the underlying motivations that drive lenders to extend loans to borrowers. In addition, it is also useful to understand the decision-making processes of lenders, as it provides insight into how (unintentional) discrimination can occur in the decision-making process.

4.4.5.2.1 The motivation of lenders

While borrower motivations are relatively clear, understanding lender motivation is also important for the long-term success of a platform, especially given the typical shortage of lenders relative to borrowers seeking funding (Prystav, 2016).

Lenz (2016) discovered that P2P lenders are motivated to invest only when the financial returns from using platforms compensate for the additional risk taken and surpass other investment opportunities of similar risk. Many P2P lending platforms have met this requirement, with lenders receiving a net risk-adjusted return ranging from 1% to 10% per year, depending on factors such as the chosen platform, risk category, and investment maturity. Additionally, A2 provided further insight by stating:

> *'Banks provide little risks but there is almost no return, at the same time, stocks can provide better returns but have much higher risks. So P2P solutions give me a flexible option to get a good return for a decent amount of risk, when choosing the right trustworthy requests'*

This comparison with the bank is also reflected in the research of Lenz (2016) who stated that lenders often weigh the return of P2P loans against the interest received on their bank account deposits. However, the author argued that this comparison does not quite hold for two reasons. The first is that deposits in bank accounts are liquid assets as they can normally be withdrawn daily. P2P lending, on the other hand, are unsecured and not fungible or liquid. This implies that lenders should be prepared to commit their capital for a longer period and hold the investment until maturity, without any guarantee of an exit option. Secondly, the EU Directive on Deposit Guarantee Schemes ensures that deposits in bank accounts with EU banks are protected by EU member states up to a limit of €100,000 per person per bank. There is no such legal guarantee with P2P lending.

4.4.5.2.2 Discrimination based on the provided information

A3 pointed out that most of the research on discrimination in P2P lending concerns human biases, particularly how people treat others differently based on the information provided. This was also found during the literature review.

Unlike traditional lending models that rely on banks as intermediaries for delegated monitoring, P2P lending often entrusts individual lenders with processing borrowers' financial and personal information to determine whether to provide loans (Prystav, 2016). One significant challenge that emerges in P2P lending markets is the issue of information asymmetry. This refers to a scenario where the lender lacks the same level of knowledge as the borrower regarding their creditworthiness (Emekter et al., 2015). As a result, lenders face a disadvantage as borrowers may not disclose all pertinent information, leading to an uneven distribution of information. This can impede the lender's ability to make an informed decision and potentially result in a suboptimal allocation of funds.

However, an important issue that arises is information asymmetry. In this context, it refers to a situation in which the lender lacks the same knowledge as the borrower, especially about the borrower's credibility (Emekter et al., 2015). This creates a disadvantage for lenders because borrowers often do not disclose all relevant information, creating an imbalance in information access. This issue is particularly challenging in P2P lending because lenders often lack access to all the necessary documents and mechanisms that banks typically have (Emekter et al., 2015). For example, traditional banks rely on collateral, certified accounts, regular reporting, and even the presence of the board of directors to enhance trust in borrowers. However, in P2P lending, lenders must judge whether they deem a borrower trustworthy based on the information they receive.

The availability and utilization of different types of information in the lending process, including both hard and soft information, can significantly impact loan funding success and potentially lead to discriminatory outcomes. Hard information typically consists of standard banking variables obtained through credit checks, such as delinquency history, credit lines, home ownership status, debt, inquiries, public records, and income (Iyer et al., 2016). Qiu et al. (2012) found that loan characteristics, including loan amount, maximum acceptable interest rate, and loan period, also play a significant role in the success of loan funding.

Soft information, on the other hand, refers to subjective, non-financial, and often harder to verify information provided by borrowers (Iyer et al., 2016). This includes personal background, descriptions of financial situations, reasons for needing a loan, income and expenditure flows, education, profession, qualifications, social media activity, race, gender, and age (Iyer et al., 2016; Sukmaningsih, 2018). Galak et al. (2019) found that lenders prefer individual borrowers over groups and people who they are friends with. Additionally, lenders prefer borrowers who share similarities in terms of gender, job, and first name initial. The use of soft information, particularly related to a borrower's appearance or personal characteristics, such as ethnicity, gender, and attractiveness, has been observed to influence lending decisions (Pope & Sydnor, 2011; Ravina, 2019). This indicates that the soft information provided can potentially lead to subjective judgments from lenders influenced by socio-physiological factors such as stereotypes.

Duarte et al. (2012) found that borrowers perceived as more trustworthy are more likely to receive loans and lower interest rates compared to those perceived as less trustworthy. Factors such as gender and ethnicity were examined in determining trustworthiness. The authors provided two potential explanations for this observation: reputational capital indicated by a person's appearance, which is difficult to manipulate, and a shared biological foundation that shapes a trustworthy appearance. The study by Cesarini et al. (2008) supports the notion that trustworthiness is partly influenced by genetics. Furthermore, research in psychology indicates that people often develop expectations and biases based on judgments made from appearance (Duarte et al., 2012). For example, Ravina (2019) demonstrated that an above-average beauty appearance increased the likelihood of obtaining a loan, with a stronger effect observed for women. However, Pope & Sydnor (2011) revealed negative effects, such as lower loan probability and higher interest rate requirements for borrowers without pictures or pictures of blacks, elderly individuals, or unhappy people.

Prystav (2016) investigated taste-based discrimination and found that loan applicants seeking funds for non-essential purposes, such as vacations or renovations, were less likely to receive support from lenders. The disclosure of the loan's purpose, coupled with other factors like spelling and grammar errors, negatively influenced lenders' perceptions. Although the study did not specifically explore discrimination factors, indirect discrimination based on educational level and race can be inferred. Projects with poorly written loan requests, often associated with lower education levels or non-native language proficiency, tend to receive less funding, indirectly affecting individuals from different ethnic backgrounds. Prystav (2016) also discovered that lenders generally overlooked projects with poor credit ratings, except when the loan's purpose was related to education or self-employment. This aligns with the study by Mach et al. (2014), which reported a higher probability of funding for small business loans, suggesting that lenders believe these loan obligations will be fulfilled. However, prioritizing these specific loan types may unintentionally exclude individuals with limited access to business opportunities or higher education.

The findings presented by Gonzalez & Loureiro (2014) highlight the existence of statistical-based discrimination in lending. They found that borrowers perceived as younger are allocated lower loans due to the belief that they have less experience and competence, even though this belief may not be true for every individual in the age group. Additionally, borrowers may provide inaccurate information or tailor their information to obtain more favourable loan terms. Duarte et al. (2012) discovered that homeowners, college graduates, and individuals perceived as trustworthy have a higher likelihood of receiving loans. However, it is important to note that individuals within these groups may not always have better credit scores compared to those outside these groups. Furthermore, Prystav (2016) identified herding behaviour among lenders when little to no information is available. Lenders tend to rely on the decisions of other lenders and choose to fund projects that have received higher previous funding. This behaviour can contribute to statistical-based discrimination.

This highlights the connection between the data output factor and lenders, wherein lenders can potentially contribute to discriminatory outcomes based on the type of information provided by the P2P lending platform.

4.4.6 User Experience Domain

This section provides an overview of the various factors and influences within the user experience domain.

4.4.6.1 User interface

A7 explained that companies do not feel the same obligation as governments to ensure equal accessibility for all citizens. Companies are not motivated to prioritize spending money on some user groups that are very small and fragmented because they are not as profitable as their main customer groups. However, A7 noted that the issue is not just related to interface design, but rather the underlying challenge of making accessibility a priority when it can impact timelines and budgets. Despite this, A7 emphasized the existence of guidelines for designing accessible interfaces. Moreover, this is not only an issue of financial organisations but of all organisations that have a digital platform.

A13 noted that linguistic minorities could face discrimination that can be indirectly linked to ethnicity. For instance, if a platform is only available in Dutch, it can create barriers and exclude non-native speakers. This means that people of certain ethnic backgrounds do not always have the same access to these platforms as others with similar characteristics, solely due to their language barrier. Cardinal et al. (2020) discuss that designers do recognize the importance of creating systems that are not limited to standardized languages such as American English due to globalization and increased migration. Despite this recognition, translation and localization work for these systems are frequently outsourced to third parties after the realization of the design or treated as a separate phase of the design process. As a result, the needs of users that are not fluent in the platform language are often not taken into account until after the system has already been designed. This can lead to designs that discriminate against communities by limiting their language access (Cardinal et al., 2020).

A7 and A11 discussed that older individuals often have different needs and abilities when it comes to using digital platforms. Many services including P2P lending are being provided digitally, with some being intended for the general population, while others target specific age groups, such as seniors, youths, or students. Studies have highlighted the challenges faced by older adults when accessing technology, including physiological, cognitive, and emotional difficulties (Manor & Herscovici, 2021). Manor & Herscovici (2021) discussed that accessibility features for people with disabilities and language accommodations for non-native speakers are not specially tailored to the needs of older individuals. In addition, designers are not always aware of the unique difficulties and needs of older users and do not always consider them as a target group, indicating implicit ageism. Ageism is a negative attitude or behaviour towards people based solely on their age and can be considered as a form of discrimination, where individuals are judged based on age-based stereotypes (Manor & Herscovici, 2021). These individuals can be seen as weak, dependent, forgetful, and more, which lead to age-based prejudice. According to Iversen et al. (2012), ageism consists of three components: cognitive (stereotyping), emotional (prejudice) and behavioural (discrimination). The exclusion of older adults from the information society are partly due to stereotypes that portray them as unwilling or unable to learn how to use new technologies. This means that older adults face unequal access to P2P lending platforms when they are not designed with their specific capabilities in mind, making it difficult for them to get loans through these platforms.

4.4.7 Business Domain

This section provides an overview of the various factors and influences within the business domain.

4.4.7.1 Process

Numerous interviewees have noted that biases can creep in at various stages of the process, from data collection to algorithm design to human decision making, making them difficult to identify and address (A3, A8, A11, A12). The reason behind this difficulty is that biases are often unconscious, making it challenging to pinpoint the stage in the process where bias has occurred (A11). In addition, decisions are often made without clear documentation, which over time makes it difficult to explain the reasoning behind a discriminatory decision (A8). This lack of transparency can make it difficult to identify where biases may have influenced the outcomes. Additionally, the distribution of responsibilities among different teams involved in the lending process can create ambiguity regarding the decisions and actions taken by other teams, further contributing to the lack of transparency and accountability.

It is found in the literature that discrimination can occur at three main stages during the contracting process for financial products (Garz et al., 2021). Firstly, the customer acquisition or discouragement stage is the initial phase where institutions try to attract or deter potential customers. Decisions about factors such as advertising strategies and customer service quality can disproportionately affect certain groups. Fairlie et al. (2021) found that black firm owners are more likely to be discouraged to borrow, even if they are considered creditworthy. They also indicated that they do not apply for loans as they are afraid to be rejected. A4 also noted this fear of rejection among borrowers and raised an interesting point about the exclusion of certain individuals through the financial products offered by organizations. They illustrated their point by highlighting that:

> *"Financial products that are structured in certain ways can also exclude … we have a million Muslims in the Netherlands who would like to finance Halal, where no taxes can be paid. No financial organization in the Netherlands offers this kind of product and that is very strange."*

This example demonstrates that failure to consider the needs of all customers can lead to the exclusion of specific groups from the beginning. This type of exclusion through product offerings was also mentioned by A7. This situation arises because organizations often prioritize their main target groups, which are typically those that bring in the most return on investment and more mature fragments of the markets.

Secondly, during the contracting stage, discrimination could happen as loan application need to be evaluated, approved, or rejected (Garz et al., 2021). This can involve offering different contract terms, interest rates, or loan amounts to applicants based on their protected characteristics (soft information), rather than their individual qualifications (hard information). In the context of P2P lending, the terms and interest rates are often calculated by an algorithm and lenders are responsible for deciding whether to lend, to which borrowers and the loan amount based on the information provided. Lastly, discrimination can continue in the post-contracting stage, even after a contract has been established (Garz et al., 2021). In this stage, unequal treatment could arise in how financial organisations manage customer relationships. These findings are aligned with the study by Atkins et al. (2022) who also found that financial organizations can discriminate against potential borrowers through at least four different mechanisms: refusing a loan application, discouraging borrowers, demanding a higher price and granting smaller amounts of credit.

4.4.7.2 Motivation and target customer

As mentioned above, companies tend to prioritize issues based on financial incentives and regulatory burdens. This often results in companies focusing on the most mature group of people. According to A7:

> *"It is prioritization, like what seems to be the biggest return is done first … organisations are not motivated to tailor their services to every customer but mainly to more mature fragments of the market."*

This prioritization can lead to certain groups or topics being neglected, especially those that are not considered financially lucrative or do not belong to the current regulatory burden.

4.4.7.3 Accountability

An interesting factor in P2P lending is accountability which refers to the capacity and willingness of individuals or organisations to recognize and accept their responsibilities, provide explanations, and take blame for their failures or mistakes. The absence of accountability can lead to several potential consequences such as the tendency to shift blame and the lack of explainability. When individuals do not perceive themselves as accountable, they are less likely to provide justifications for their actions. Since P2P lending companies see themselves as intermediaries and lenders deciding whether to invest in a given loan, the algorithms determine the interest rates and loan amounts. This can lead the companies to easily shift the responsibility to the lenders and algorithms. A3 highlighted this issue stating that:

> *"If they do not have an accountability mechanism, they can just blame it on the algorithm because it is not intended unless they have specific testing procedures, policies and processes in place at the organizational level. Right now, it's kind of discretionary in terms of, are you going to test for this or not?"*

A12 stressed the importance of P2P operators taking accountability as they are involved in assessing borrowers' creditworthiness, setting loan criteria, setting interest rates, and providing this information to lenders to make informed decisions. Moreover, Prystav (2016) found that P2P lending operators have the power to modify the information accessible to potential lenders, such as credit scores, loan duration, funding progress, interest rates, and personal details. This shows the potential impact operators can have on the lending decisions of lenders.

4.4.7.4 Employees

Discrimination in (P2P) lending can also arise from the employees involved in various aspects of the lending process. For instance, studies have shown that the financial industry is predominantly male, which may result in unintentional cultural prejudice towards female business owners (Orhan, 2019). This is also noted by A4. However, P2P lending platforms differ from traditional financial institutions. P2P lending platforms facilitate direct lending between the borrower and lender and many processes are also automated by technologies such as the decision-making regarding loan amount. Morse et al. (2021). discussed that the use of algorithms in decision-making roles can replace human discretion, reducing taste-based discrimination.

While technology can seem like a neutral tool, it does not always eliminate human discretion, as humans are still involved in various aspects of technology (Morse & Pence, 2021a). As discussed before, P2P lending platforms are information portals consisting of both hard and soft information for lenders to decide whether to lend for a particular loan application. This places the responsibility on lenders and leads to taste-based biases among lenders (Morse & Pence, 2021a). In addition, coding and data inputs can lead to discrimination, as algorithms are ultimately designed by humans, even if AI handles the mechanics (Morse & Pence, 2021a). In particular, the programmers responsible for setting up the optimization problem, classifying the data, and choosing the training data sets can influence the output of the algorithm. Training datasets can cause discrimination as they embed existing prejudices in society or as they are not representative of the broader population (Morse & Pence, 2021a). This is discussed in more detail in the Data Domain section.

However, Chander (2016) argued that programmers do not intentionally manipulate algorithms leading to these discriminatory outcomes for several reasons. First, societal discrimination is often unconscious and less likely to be incorporated into automated algorithms, requiring a conscious understanding of what is being sought, compared to human decision-makers. The programming process requires explicit instructions and documentation about what the program does, making unconscious discrimination less likely to manifest itself in programming. Second, even if programmers intend to discriminate, the coding itself is likely to deter them from coding such discrimination, as there is a chance that the hard-coded discrimination will later be exposed by hackers or insiders. Furthermore, since writing, reviewing, and debugging code involve teams of programmers, it is likely that deliberate coding discrimination will require the cooperation of multiple programmers, which is a difficult task. Finally, concerns about the black box of human decision-making have driven some companies to turn to algorithms as a solution and not as an objective to discriminate.

4.4.8 Data Domain

This section provides an overview of the various factors and influences within the data domain.

4.4.8.1 Input data

A10 provided an overview of the standard process for building credit risk models used by financial organizations, particularly banks in the Netherlands. Although there are a few pre-existing credit risk models available from the European Central Bank, these models are not widely used due to their unattractive features. Instead, most financial organizations generally prefer to create their own models, which usually involves gathering historical data for the past 10 to 15 years on all their clients. The next step is to analyse the performance of these clients and identify the parameters that are correlated with their repayment behaviour. These parameters are then used as indicators to forecast the repayment behaviour of new clients in the future. A10 summarized this process by stating that this process involves using a large amount of historical data to identify patterns and develop a predictive model. However, A11 raised an important concern about the representativeness of the historical data used in these models, stating that:

> *"The transaction data of the customers represents the population and the demographics of the people who bank with that particular bank. This does not necessarily mean that it represents the demographics of the entire population of the Netherlands."*

However, they acknowledged that given the size of the bank, there may be some correlation between the two demographics, and this is more of an issue for smaller financial organizations such as P2P lending operators. However, A12 highlighted that P2P lending companies in the Netherlands might encounter difficulties when trying to accurately assess the creditworthiness of their customers, particularly in the early phases of operation. This is due to the lack of access to historical data from banks and the absence of credit agencies like those in the United States. Moreover, A3 stated an interesting point by stating that one of the biggest issues in lending in general is the unknown unknowns, thus if you do not give a loan to someone, you will never know if they would have repaid it. They explained by stating:

> *"If you build an algorithm based on past approvals, you're only going to give loans to similar people and that naturally excludes previously just excluded populations and marginalized populations."*

Input data consisting of proxies have been mentioned by several interviewees (A1, A2, A6, A8, A10, A12) as a factor that can contribute to discrimination. There are various input data that can be used to feed models, including behavioural data. As A12 pointed out, behavioural data may contain hidden correlations that can be learned by algorithms. For example, financial organizations commonly use transaction data to capture details such as an individual's payment locations, amounts spent, and subscription types. However, such data can potentially lead to discrimination in lending practices. To illustrate this point, A12 gave an example of how a person's ethnicity can be correlated with his literacy level. People who cannot read or write are often immigrants or individuals who have fallen outside of the education system. These individuals may have different payment behaviours than those who can read and write, leading to hidden correlations within the behavioural data. Consequently, this hidden correlation can lead to unintended discrimination in lending decisions. However, an additional challenge emerges as the ethnicity-related data is subject to privacy protection, making it difficult to verify whether it is actually linked to observed pattern. This issue will be further explored in the Challenges Chapter.

4.4.8.2 Data process

The data used as input for a model is typically pre-processed before being fed into the model. However, this preprocessing stage can introduce discrimination that is more problematic than other stages. A8 elaborated on this issue, stating that:

> *"Most invisible and the worst one is in the preprocessing because it will never show up later if it is already excluded at that point in time. It is the least transparent and least easy to bring to the surface ... while if the discrimination happens at the end of the process in the visualization, it is more obvious or easy to measure."*

In addition, A3 argued that preprocessing should only be used in very specific cases as it can harm a model's performance, explainability and robustness. They noted that preprocessing can essentially screw up the input dataset, which can have negative consequences for the customers of the model, especially in sensitive applications such as credit risk models.

4.4.8.3 Output data

In P2P lending context, operators of P2P platforms can determine which information is disclosed in loan applications, that could potentially impact the investment choices of lenders (Prystav, 2016). This data provides insights into risk levels and additional borrower information that can influence lenders' lending decisions. However, if the data is inaccurate or misrepresented, it could worsen discrimination instead of reducing it. A3 also highlighted the role of biases among lenders, which poses a discrimination issue in P2P lending. This shows the role of output data to discrimination in lending.

4.4.9 Technology Domain

This section provides an overview of the various factors and influences within the technology domain.

The findings from the interviews strongly indicated that the root cause of algorithmic decision-making that leads to discriminatory outcomes is primarily related to human, the input data and parameters utilized, rather than the algorithm itself. As highlighted by A3:

> *"An algorithm cannot solve a non-algorithmic problem and everything around that."*

These algorithms are designed as mathematical models that rely on predetermined rules and parameters to process data and generate outputs. However, these rules and parameters are established and defined by humans who often have their own biases and preconceptions. Data inputs, such as incomplete and unrepresentative data, can also impact the algorithmic output and ultimately result in discriminatory outcomes. These factors of people interaction and data have already been explained in the previous sections. The next section examines how discrimination can arise from the subfactor of parameters.

4.4.9.1 Parameters used in models

A10 emphasized that assessing one's creditworthiness is a complicated procedure in which many parameters are taken into account. In the Netherlands, financial organizations commonly rely on parameters such as the BKR, income, and employment to assess creditworthiness. However, A10 expressed concern about the limitations of such a model, stating that:

> *"It could be the case that certain individuals just end up very unlucky due to the parameters that are being considered, resulting in sort of false negatives."*

This concern was illustrated by an example shared by A8, who explained that financial organizations could unintentionally exclude specific customer groups due to the parameters set within their risk models. A8 described a scenario where a small company seeking a mortgage for a workspace smaller than 30 square meters would face challenges due to the model parameters, which were designed to only consider workplaces larger than this size. The exclusion of certain customer groups was unintentional, because financial organizations were not informed that they were missing out on potential customers.

Moreover, BKR has a monopoly on the Dutch credit market. As A10 pointed out, this can lead to situations where individuals are falsely denied loans or other financial opportunities due to negative BKR codes that do not accurately reflect their current creditworthiness. This issue came up recently in an article shared by A10 describing a case where a person had a BKR registration due to past financial issues that had already been resolved, but still struggled to get approved for a mortgage.

4.5 Conclusion

The chapter aimed to identify the factors contributing to financial discrimination in the lending system. To make the concept of discrimination more applicable in the context of financial lending, specific dimensions were identified and defined: differential treatment, differential outcome, unequal access, and lack of transparency.

- Differential treatment refers to intentional discrimination based on protected characteristics.
- Differential outcome occurs when there is no explicit intent to discriminate, but the treatment results in unequal lending outcomes.
- Unequal access refers to exclusion from loans due to platform design, occurring even before the loan process starts.
- Lack of transparency refers to financial organizations not providing explanations for loan outcomes, making it challenging to prove discrimination.

The chapter identified various factors that can contribute to one or more dimensions of discrimination. These factors shed light on the domains in which discrimination can arise and how they interact within the lending context. The factors were categorized into six domains: regulatory, human interaction, UX, business, data, and technical, with some factors overlapping across multiple domains.

Regulatory domain

Strict regulations, particularly regarding maximum interest rates, make it unprofitable for P2P lending platforms to operate in the Netherlands. This limitation has hindered the growth of P2P lending in the Netherlands compared to countries without such restrictions. Organizations prioritize tasks based on regulatory demands, allocating resources to urgent compliance issues. This can lead to initiatives towards non-discrimination being deprioritized. Moreover, the lack of specific guidelines and legal weight behind addressing discrimination can hinder efforts to effectively tackle discrimination in financial organizations.

Human interaction domain

Social psychological factors, including stereotypes, motivation and culture can give rise to biases that contribute to discrimination. For instance, in the Netherlands, male-dominated industries often favor masculine traits in defining entrepreneurial success, which can disadvantage women and marginalized groups. Lack of awareness among lenders is another contributing factor. While many lenders believe they offer equal opportunities to all borrowers, closer examination often reveals biases in their investment practices. This indicates that lenders may not be fully conscious of their own biases and discriminatory behaviors.

User experience domain

Companies often prioritize user groups that generate higher profits, inadvertently overlooking the abilities and needs of smaller or fragmented user groups during the design of UI. This exclusion can lead to barriers and unequal access to loans for certain individuals. Language barriers can also lead to exclusion, as platforms that are only available in a specific language exclude non-native speakers, particularly affecting linguistic minorities.

Business Domain

Throughout the lending process, biases can emerge at various stages, making them challenging to identify and address. The reason behind this difficulty is that biases are often unconscious, and decisions are often made without clear documentation, making it difficult to trace the sources of discrimination. Moreover, the distribution of responsibilities among various teams involved in the lending process can create ambiguity regarding the decisions and actions taken by others, further contributing to the lack of transparency and accountability. Moreover, companies prioritize issues based on financial incentives and regulatory burdens, which can result in certain groups or topics being neglected if they are not considered financially lucrative or do not fall within the current regulatory focus.

Data Domain

The input data used to build credit risk models often relies on historical data collected by financial organizations. However, these do not accurately represent the demographics of the entire population, leading to potential biases in the model's predictions. Certain types of data used in input can act as proxies, introducing hidden correlations that may lead to unintended discrimination in lending decisions. Additionally, the presence of unknown unknowns in lending introduces challenges as denying a loan to an individual means never discovering whether they would have been capable of repaying it. This can inadvertently result in the exclusion of certain populations from accessing loans and opportunities.

Technology Domain

Algorithmic decision-making can lead to discriminatory outcomes, but the root cause is primarily related to humans, unrepresentative input data, and parameters rather than the algorithm itself. Algorithms are mathematical models that rely on predetermined rules and parameters, which are established by humans who may have their own biases. Biases can be introduced through the design of parameters used in these models which can unintentionally exclude certain customer groups. This can result in false negatives, where individuals who are creditworthy are wrongly denied loans due to the parameters considered.

After understanding the factors and domains that can contribute to discrimination in the financial lending system, the next chapter will focus on identifying and understanding the challenges associated with realizing a non-discriminatory financial lending system. It is important to note that the subsequent chapters will not consider the UX domain since accessibility challenges are prevalent across all organizations that use platforms, making it a broader issue that is not unique to financial lending. Additionally, no specific factors related to the application layer of the architecture were identified, indicating a lack of direct influences in this layer.

5 Challenges

After gaining an understanding of the dimensions, factors, and domains of discrimination, this chapter aims to identify the challenges encountered in realizing a non-discriminatory financial lending system in the Netherlands. By identifying these challenges, the aim is to recognize the obstacles and difficulties associated with this objective and subsequently develop design principles that address these challenges. To address this objective, the following sRQ will be addressed:

What are the challenges for realizing a non-discriminatory financial lending system?

First, the report presents the findings of a SLR conducted to identify challenges based on existing studies. Subsequently, the results of the semi-structured interviews are presented to assess the applicability of these challenges in the Dutch context and to identify any additional challenges experienced by the participants. The findings from both sources are then synthesized to form a comprehensive set of challenges. These challenges will serve as the basis for the next chapter, where architects will be presented with the principles derived from these challenges and tasked with addressing them.

5.1 Systematic literature review

This SLR focused on investigating the challenges related to realizing a non-discriminatory P2P lending system in the Netherlands. The search was conducted using Scopus with a query that included keywords related to P2P lending and discrimination. The query used was:
TITLE-ABS-KEY (("DISCRIMINAT*" OR "BIAS" OR "FAIR*" OR "EQUAL*") AND ("P2P LEND*" OR "PEER-TO-PEER LEND*"))

This query resulted in 91 sources. The manual filtering process was then applied, starting with the evaluation of abstracts for relevance. 40 sources were filtered out as they were deemed not relevant. These included sources that mentioned discrimination as the ability to distinguish or as a discriminative feature, as well as papers that used the term 'equal' in a different context. The remaining 51 sources underwent a full-text assessment for eligibility. Out of these, 13 sources were excluded as they were predominantly focused on technical details, such as machine learning algorithm implementation, or examined regulations and laws outside of Europe. Ultimately, a total of 38 sources were selected based on their explicit focus on P2P lending discrimination. Through the snowballing method, three additional sources were added, resulting in a total of 41 sources. Among these sources, relevant challenges were identified in 30 of them, forming the basis for the challenges identified in this literature review. The final list of sources can be found in Appendix F.

A full-text analysis is performed on the 30 relevant sources to identify potential challenges associated with discrimination in P2P lending. This analysis resulted in the identification of 50 challenges, which were then consolidated into 26 unique challenges. These challenges are further categorized into domains from the previous chapter. Each challenge is allocated to a specific domain, although it is important to note that some challenges may have relevance across multiple domains. The allocation is based on the most direct association with a particular domain. Table 3 presents the 26 challenges, including their ID, Domain ID, Challenge Name, Short Description, and Source ID. This ensures that the challenges can be easily traced back to their respective sources and facilitates the merging of challenges from interviews into a final set.

As indicated in Table 4, the challenges are categorized into different domains. The data domain has the highest number of challenges, with a total of 10. This is consistent with the common theme identified in the literature regarding potential discrimination based on borrower and platform information. The technology and human interaction domains each have 5 challenges, highlighting the significance of technological and human factors in addressing discrimination in P2P lending. The business domain encompasses 4 challenges. Lastly, the regulatory domain has 2 challenges, which may be attributed to the exclusion of regulations and laws outside of Europe during the selection process.

Table 3: All 26 identified challenges through literature review categorized by domain

Unique ID	Domain ID	Challenge Name	Short description	Source ID
C-Li1	Bus01	Difficult identification and measurement of discrimination impact	Loan discrimination can happen in different ways and might be platform-specific, making assessment of discrimination challenging due to varying methodologies used to measure it, and potential factors behind bias may not always be fully accounted for in the evaluation	P18, P26, P3
C-Li2	Bus02	Human involvement introduces bias and discrimination	Human involvement in the selection, design, and coding of algorithms can result in the encoding of prejudice, biases, and discrimination into the system	P1, P2
C-Li3	Bus03	Profit maximization can lead to statistical discrimination	Even when purely motivated by maximizing profit, one might be applying statistical discrimination by profiling information that led to discrimination	P2
C-Li4	Bus04	Group solutions might not help individuals	Solutions to mitigate discrimination against a group may not effectively address this issue for certain individual members of that group	P12
C-Li5	Data01	Personal data can be unavailable or faulty	Discrimination is difficult to identify if personal data is not available or faulty	P1, P4
C-Li6	Data02	Long-standing biases can be present in data	Input data can contain long-standing biases related to gender, race, geography, economics, and other factors that translate to the model and output	P1, P2
C-Li7	Data03	Unintentional revelation of sensitive data	Input data used by algorithms may unintentionally reveal sensitive personal information leading to discrimination, which may exist even without direct personal information availability, as it may be deducted indirectly	P1, P3
C-Li8	Data04	Ineffective sensitive data removal	Overwriting sensitive information as an input can still result in biased outcomes	P1
C-Li9	Data05	Input data can be skewed	Input data could be based on a sample size that is small or skewed towards certain groups that are not representative of the entire population	P1
C-Li10	Data06	Unclear relation to sensitive data	It is difficult to check whether a variable is associated and if so, to what degree with discriminatory attributes, e.g., neighbourhood can be correlated with certain ethnicities	P1
C-Li11	Data07	Balancing accuracy and usage of personal data	Variables that are closely related to identifying minorities are often avoided in models to minimize the chance of unintentional discrimination at the expense of accuracy	P4
C-Li12	Data08	Sole reliance on specific data	Credit scoring systems often rely solely on data from one specific group and do not factor in information from other groups, leading to bias	P9
C-Li13	Data09	Lack of available historical data	P2P platforms often lack sufficient financial history to get an accurate view of the risks related to a borrower, leading to usage of alternative methods to assess creditworthiness	P7
C-Li14	Data10	Exclusion due to variables used	By embedding certain selective variables into the credit-scoring process, certain groups might get excluded already	P2
C-Li15	Hum01	Lender's traditional and behavior biases	Lenders face a range of biases when selecting borrowers, including biases related to demographics, geography, status, attractiveness, and beliefs, which can result in differential treatment of borrowers and may not necessarily reflect the borrowers' actual creditworthiness	P14, P15, P16, P17, P19, P21, P22, P23, P24, P25, P27, P29, P30, P5, P6
C-Li16	Hum02	Borrowers becoming discouraged to apply	Certain groups might be less likely to apply for a loan due to lack of self-confidence, perception of discrimination, perception of lower creditworthiness and/or fear of rejection	P1, P11
C-Li17	Hum03	Proxies can be inaccurate and bring biases	Lenders might use proxies or other methods to gather information on borrowers, and make decisions based on unobservable factors that are difficult to measure	P13, P28

C-Li18	Hum04	Providing additional information can backfire	Providing lenders with additional information to reduce bias might not only be ineffective but even worsen it as they might use it to validate existing biases and form even stronger opinions	P20
C-Li19	Hum05	Balancing information presentation	The way how information is presented to lenders can influence their decision behaviours and therewith reduce or increase bias	P6
C-Li20	Reg01	Uncertainty surrounding GDPR	There exists lots of uncertainty around how the GDPR will be applied and interpreted, particularly when it comes to statistical discrimination based on special-category data	P2
C-Li21	Reg02	Difficulty of implementing non-discrimination laws	Non-discrimination laws are complicated to implement in the EU due to difficult adherence of platforms to national laws of each member state and language barriers	P10
C-Li22	Tech01	Personalization can lead to statistical discrimination	A potential consequence of using personalized algorithms based on big data is the possibility of statistical discrimination, potentially leading to groups being disadvantaged compared to others	P3, P8
C-Li23	Tech02	Difficulty of observing and assessing algorithmic bias	Observing learned biases in algorithms and assessing their fairness is challenging due to limited external visibility and the contextual nature of what is considered fair or biased	P1
C-Li24	Tech03	Keeping algorithms up to date with societal values	Operational algorithms need to be revised and updated with the evolving context of bias and discrimination within society and cultural values	P1
C-Li25	Tech04	Increasing accuracy can introduce bias	Methods to increase accuracy of risk profiles of borrowers can introduce bias and lead to discrimination	P7
C-Li26	Tech05	Solutions unintentionally introducing other biases/drawbacks	Solutions introduced to reduce bias might unintentionally introduce other biases or drawbacks to borrowers	P5

Table 4: Number of unique challenges identified for each dimension

Domain	Code abbreviation	#Unique challenges
Data	Data	10
Technology	Tech	5
Human Interaction	Hum	5
Business	Bus	4
Regulatory	Reg	2
Sum		**26**

Table 5 offers more detailed information regarding the target countries of the relevant sources. The majority of research primarily focuses on P2P platforms active in the USA and China, while European platforms are underrepresented. The reasons for the less developed state of P2P lending in Europe are discussed in Chapter 4. The data derived from these studies will be utilized to assess potential discriminatory practices within P2P lending platforms. Interviews will provide additional insights into the applicability of these findings to the Dutch market.

Table 5: Target country of research

Country	Unique papers
USA	12
China	9
Developing countries	4
Not applicable (NA)	2
Korea	1
Israel	1
Germany	1

5.2 Semi-structured interviews

The data from the 13 interviews conducted was analysed using Atlas.ti to derive the key challenges for realizing a non-discriminatory lending system. The coding process was carried out in two stages: open coding and merging codes.

During the open coding stage, a total of 82 codes were identified. However, it was found that 10 of these codes were duplicates from the same interview, resulting in 72 unique codes (see Appendix F, Table F2). Each of these codes have a unique ID such as C-In1 to ensure these can be easily traced back to their respective sources. It is important to note that while the interviews specifically asked about challenges, most of the responses provided by the interviewees were also explanation of the challenge or problems rather than challenges. For instance, A6, A7, and A13 mentioned the challenge of organizations being reactive rather than proactive in addressing discrimination, only taking action when there is a complaint, reputational risk, or legal requirements (C-In11, C-In12, Cin13). The reason behind the challenge was further explained by A11, who highlighted time constraints and compliance-driven changes in financial organizations that often result in anti-discrimination initiatives not being prioritized or constantly postponed (C-In3). Additionally, A1 emphasized the problem that discrimination has not received the same level of attention and prioritization as other important issues like privacy (C-In4). These problems were then merged and combined into a challenge titled "Organizations tend to react to incidents of discrimination rather than being proactive in preventing them".

Through the process of combining and merging the identified codes, a final set of 13 challenges was derived (Table 6), representing the key issues associated with achieving a non-discriminatory lending system.

5.2.1 Synthesis

It is observed that the challenges identified from the literature review can be incorporated into the final set of challenges obtained from the interviews. However, there are more challenges identified from the interviews compared to the literature review. This difference can be attributed to the specific focus of the literature search query on P2P lending, while most of the challenges discussed during interviews were about financial lending systems in general.

One notable finding was that the literature emphasized the potential for lenders to discriminate based on the information provided by borrowers and the platform. This was a recurring theme in the literature, but only one A10 specifically mentioned this challenge. This can be attributed to the fact that the interviewees were all based in the Netherlands, where the P2P lending market is relatively limited compared to other countries. As a result, their direct experience with P2P lending platforms might have been more limited.

In the existing literature, there is no direct mention of humans being unaware of their own bias. However, this concept is closely related to C-1. Additionally, it is not just limited to lenders who may be unaware of their bias, but also extends to employees within financial organizations. It is challenging for individuals to recognize their own prejudices, particularly in a society where certain biases are normalized and not everyone is equally affected by them. In addition to the challenges identified in the literature, there are other significant factors that contribute to the complexity of addressing bias in lending. One such challenge is the presence of unknown factors that can lead to biased outcomes. Detecting and addressing bias sources in a model before its implementation is a broader issue that goes beyond the specific context of P2P lending. The presence of unknown underlying factors makes it difficult to anticipate and prevent all instances of bias.

Another challenge that emerged during interviews is the tendency of financial organizations to evade responsibility and accountability for discriminatory outcomes. Currently, these organizations lack mechanisms to hold themselves accountable for the outcomes resulting from algorithmic decisions. This often allows them to attribute unintended outcomes solely to the algorithm, thereby evading responsibility. Additionally, financial organizations often lack transparency and explainability in their algorithms. In the case of P2P lending platforms, the ultimate responsibility for lending decisions lies with individual lenders, potentially allowing the platform to evade accountability for any discriminatory outcomes.

Table 6: Final 13 challenges for realizing a non-discriminatory financial lending system

ID	Challenge	Domain(s)	Interviewees	Literature	#Interviewees	#Literature	#Total citations	Underlying challenges
C-1	Lenders can discriminate based on the provided information by borrowers and the platform	Human Interaction	A10	P14, P15, P16, P17, P19, P21, P22, P23, P24, P25, P27, P29, P30, P5, P6, P6	1	15	16	C-In1, C-Li15, C-Li19
C-2	Organisations tend to react to incidents of discrimination rather than being proactive in preventing them	Business	A1, A11, A12, A13, A3, A6, A7	P2, P3, P8	7	3	10	C-In10, C-In11, C-In12, C-In13, C-In14, C-In2, C-In3, C-In4, C-In5, C-In6, C-In7, C-In8, C-In9, C-Li22, C-Li3
C-3	Indirect discrimination through proxies is difficult to detect	Data	A10, A11, A12	P3, P1, P4, P13, P28, P2, P4	3	6	9	C-In15, C-In16, C-In17, C-In18, C-In19, C-In20, C-In21, C-In22, C-In23, C-Li11, C-Li17, C-Li20, C-Li5, C-Li7, C-Li8
C-4	Difficulty in discovering and tracing where discrimination or bias occurred throughout the process	Business	A11, A12, A3, A8	P1, P2, P18, P26, P3	4	5	9	C-In24, C-In25, C-In26, C-In27, C-In28, C-In29, C-In30, C-In31, C-In32, C-Li1, C-Li2, C-Li23
C-5	Difficulty for borrowers to prove that there is discrimination	Human Interaction	A1, A13, A3, A5, A6	P1, P11	5	2	7	C-In33, C-In34, C-In35, C-In36, C-In37, C-In38, C-In39, C-Li16
C-6	There is a lack of concrete guidance on how to ensure that a system or	Regulatory	A1, A11, A12, A6, A7	P10	5	1	6	C-In40, C-In41, C-In42, C-In43, C-

	service is not discriminatory.							In44, C-In45, C-Li21
C-7	Solutions to reduce bias and discrimination can have unintended negative effects	Business, Data & Technical	A12, A3	P12, P20, P5, P7	2	4	6	C-In46, C-In47, C-Li18, C-Li25, C-Li26, C-Li4
C-8	Human involvement introduces biases throughout the process	Human Interaction & Business	A10, A12, A3	P1, P2	3	2	5	C-In48, C-In49, C-In50, C-In51, C-Li6
C-9	Humans are not aware of their own biases	Human Interaction	A11, A12, A2, A4, A8		5	0	5	C-In52, C-In53, C-In54, C-In55, C-In56, C-In57
C-10	Bias can be introduced through data in many ways	Data	A11, A3	P1, P2, P7, P9	2	3	5	C-In58, C-In59, C-In60, C-In61, C-Li10, C-Li12, C-Li13, C-Li14, C-Li9
C-11	Contextual factors can be overlooked by systems	Data & Technical	A11, A12, A4	P1	3	1	4	C-In62, C-In63, C-In64, C-In65, C-Li24
C-12	Difficult to prevent all incidents due to unknown factors	Data & Technical	A11, A13, A8		3	0	3	C-In66, C-In67, C-In68, C-In69
C-13	Financial organisations tend to avoid responsibility and accountability	Business	A12, A13, A3		3	0	3	C-In70, C-In71, C-In72

1. Lenders can discriminate based on the provided information by borrowers and the platform

Lenders can access various information about the borrower, including age, ethnicity, geographic location, and appearance. While this information is helpful in making decisions by increasing trust, it also presents a potential for discrimination. Lenders relying on such information may (un)intentionally make biased decisions, leading to unequal treatment of borrowers from certain groups.

2. Organisations tend to react to incidents of discrimination rather than being proactive in preventing them

Discrimination does not receive the same attention as other issues such as privacy. This is because financial organizations are often affected by time constraints and compliancy-driven changes, which can lead to certain initiatives not being prioritized or continually postponed. In addition, prioritization is often driven by the greatest return on investment, leaving certain groups behind. That is why organizations are reactive rather than proactive when it comes to issues such as discrimination. This means that they only act if it damages their reputation or business results if someone complains about it or if it is required by law to do so. However, it is important to note that organisations cannot react on incidents when incidents cannot be easily proven by borrowers.

3. Indirect discrimination through proxies is difficult to detect

Due to privacy regulations such as the GDPR, certain data points, such as someone's ethnicity, cannot be collected or processed. This means that when certain patterns are observed in a dataset, it is difficult to discover if it is related to ethnicity through proxies, such as zip code, as these obvious data points are filtered out.

4. Difficulty in discovering and tracing where discrimination or bias occurred throughout the process

The challenge is that when decisions are made without transparency or clear documentation, it can become difficult to explain the reasoning behind a decision that has led to a discriminatory outcome over time. This is because biases can creep into various process stages, from data collection to algorithm design to human decision making, making them difficult to identify and address. Pre-processing data can exclude or transform data in a way that it is not representative anymore. It also is often the most invisible form of discrimination. This is because if data is excluded or biased at this stage, it will never appear later in the process. Moreover, because bias is often unconscious, it is difficult to determine where in the process the bias originated.

5. Difficulty for borrowers to prove that there is discrimination

The challenge is that indirect discrimination is difficult to prove by borrowers, which makes it more difficult to address. This is because borrowers often do not complain or speak up about it, as it is a sensitive subject and there is often a lack of evidence. It is difficult to prove when an algorithm decides that the algorithm has discriminated against that borrower due to the lack of transparency. In addition, companies will always deny that they have discriminated against that borrower. They often do not take the responsibility or blame it on the algorithm, leading to certain borrowers becoming less likely to apply for a loan due to lack of self-confidence and/or fear of rejection.

6. There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory

Currently, there is a lack of generally accepted and specific guidelines or standards for implementing anti-discrimination laws in practice. Similarly, financial organizations do not have concrete guidelines for designing systems that are free from discrimination from the beginning. In contrast, principles around the topic of privacy such as PbD and the GDPR are well established and that provide concrete guidelines for organizations to follow.

7. Solutions to reduce bias and discrimination can have unintended negative effects

Solutions to mitigate discrimination might have unintended negative consequences. These may not become apparent until after they have been implemented. Even models that are designed to reduce bias can sometimes introduce new forms of bias. For instance, providing lenders with more information to help them reduce bias might end up validating the biases they already have. Additionally, providing loans to individuals who cannot afford them may seem like a solution to discrimination, but it can ultimately harm these individuals in the long run.

8. Human involvement introduce biases throughout the process

The challenge is that when working with people there will always be a subjective element involved in making decisions. For example, building a model involves subjective human judgments in determining which parameters to consider, potentially introducing biases into the model. This can be influenced by the viewpoint, stereotypes or culture of that employee involved.

9. Humans are not aware of their own biases

It is difficult for individuals to be aware of their own prejudices, especially in a society where certain biases are normalized and not everyone is equally affected by them. Biases often occur unconsciously and are shaped by an individual's point of view and background. Gender stereotypes are also rooted in the financial sector, making this issue even more challenging.

10. Bias can be introduced through data in many ways

Bias in data can arise from missing data points, incomplete data, reliance on data from specific groups, faulty data, skewed data, or by using selective variables. These might not be a good representation and therefore introduce bias. One of the causes is the unknown unknowns, such as relying on past approvals data leading to lending only to similar people. Another is the lack of historical data, requiring P2P lending platforms to use other data sources such as social media profiles and online behaviour. Furthermore, bias might be hard coded in data due to a lack of diversity in the development process.

11. Contextual factors can be overlooked by systems

Designing financial systems that take borrower's contexts into account can be difficult when cultural practices are for example not well understood. Models can label borrowers as a high-risk based solely on their spending habits, regardless of cultural context. Some cultures dictate a monthly income to their family or village, leading to above-average spending patterns that can be mistakenly interpreted as risky.

12. Difficult to prevent all incidents due to unknown factors

Detecting and addressing all sources of bias in a model in advance is difficult, as there may be unknown underlying factors contributing to bias that are not known until implementation. This also means that it can be difficult to identify all subgroups that could be affected in advance, making it impossible to guarantee the prevention of all incidents.

13. Financial organizations tend to avoid responsibility and accountability

Organizations currently lack accountability mechanisms to address discriminatory outcomes resulting from algorithms. This often leads to the possibility of blaming the algorithm for unintended outcomes. Moreover, they also often lack transparency and explainability in their algorithms. On P2P lending platforms, the responsibility for final lending decisions lies with the individual lenders, potentially allowing the platform to avoid responsibility for any discriminatory outcomes.

5.3 Conclusion

This chapter has identified and examined the challenges associated with realizing a non-discriminatory financial lending system in the Netherlands. SLR and interviews were conducted to gather information on these challenges. The SLR focused on P2P lending discrimination and resulted in 38 selected sources, while the 13 interviews provided additional insights into the challenges faced in general financial lending systems in the Netherlands. A total of 13 challenges were derived from the synthesis of the SLR and interviews:

1. Lenders can discriminate based on the provided information by borrowers and the platform
2. Organisations tend to react to incidents of discrimination rather than being proactive in preventing them
3. Indirect discrimination through proxies is difficult to detect
4. Difficulty in discovering and tracing where discrimination or bias occurred throughout the process
5. Difficulty for borrowers to prove that there is discrimination
6. There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.
7. Solutions to reduce bias and discrimination can have unintended negative effects
8. Human involvement introduce biases throughout the process
9. Humans are not aware of their own biases
10. Bias can be introduced through data in many ways
11. Contextual factors can be overlooked by systems
12. Difficult to prevent all incidents due to unknown factors
13. Financial organisations tend to avoid responsibility and accountability

The challenges identified in the SLR were consistent with those identified in the interviews, although the interviews revealed additional challenges compared to the literature. The difference can be attributed to the literature's specific focus on P2P lending, while the interviews discussed challenges in financial lending systems more broadly. Notably, while the literature emphasized the potential for lenders to discriminate based on borrower and platform information, only one interviewee specifically mentioned this challenge, possibly due to limited direct experience with P2P lending platforms in the Netherlands.

In the next chapter, a first version of design principles will be developed together with architects that addresses the 13 challenges and provide direction on how to develop a non-discriminatory financial lending system.

6 Design Principles

To address the challenges associated with realizing a non-discriminatory financial lending system, the next step involves developing design principles in collaboration with architects. This chapter aims to answer the following sRQ:

What are design principles for a non-discriminatory financial lending system?

This chapter will begin by outlining the process of deriving the design principles. Subsequently, it will present the key overarching insights obtained from interviews conducted. Following that, each individual principle will be explained. Next, an overview table will be provided, illustrating how these principles address the challenges discussed in the previous chapter. Lastly, the principles will be ranked in terms of their relevance, allowing for an exploration of any differences in the perceived importance of these principles between business and technical architects."

6.1 Approach

The design principles were derived from semi-structured interviews conducted with seven architects including 4 men and 3 women (Table 7), who were specifically asked to propose design principles that address the main challenges identified in the previous chapter. They were specifically selected based on their experience working with financial organizations (purpose sampling). These architects have a combination of practical experience from working directly for financial organizations in the past, as well as their current role as (enterprise) architecture consultants working on projects for multiple financial organizations. This combination allows them to bring both in-depth and practical insights from their direct experience, as well as a broader perspective gained from working with different companies. The architects involved in the interviews have extensive experience ranging from 5 to 25 years in the field of architecture and financial organizations, making their collective expertise and knowledge representative of architects working in financial organizations. However, it is important to acknowledge that the sample size of seven architects may have limitations in capturing the full range of perspectives and coverage of the topic. This potential limitation will be further discussed in the limitations section of the study.

The interviews were conducted online and had a duration of 60 to 80 minutes each, allowing for in-depth discussions. Transcriptions of the interviews were made and analysed. To facilitate the interviews and save time, the architects were provided with an email a week prior, explaining the interview process and providing information on the challenges they needed to address.

The architects involved in the interviews represented a mix of technical and business backgrounds. Among the sample of architects, three participants had a business background, while four had a technical background. This distinction was made to explore whether there were any differences in the principles proposed by these two groups. During the interviews, the architects expressed that there is currently a lack of existing guidance or best practices specifically addressing non-discrimination in financial organizations or other organizations. Therefore, their inputs were primarily based on their own experiences, existing guidance related to safeguarding other values such as privacy, and potential solutions. A summary of each interview can be found in Appendix H.

Given the diverse perspectives and ideas expressed during the interviews, the inputs were synthesized and formulated into design principles. An evaluation survey (Appendix G) was sent to the same group to ensure consensus and agreement among the architects regarding the formulation and content of the principles. The architects provided their feedback and suggestions through the survey, and further clarification was sought from several architects when needed through additional meetings or email. The evaluation survey itself including the adjustments, additions, or removals made to the initial principles based on the architects' feedback, can be found in Appendix I.

Table 7: Overview of the architects' background and their expertise

ID	Date	Background	Expertise
B1	12-05-2023	Business	Lead Enterprise Architecture Specialist with 10 years of experience with multiple major Dutch financial institutions
B2	15-05-2023	Business	Business architect with 5 years of experience working with Dutch banks
B3	16-05-2023	Technical	Senior IT strategy architect with 13 years of experience as enterprise architect at a large bank and insurance company
B4	16-05-2023	Technical	Senior enterprise architect with 14 years of experience advising Dutch banks and insurance companies
B5	16-05-2023	Technical	Enterprise architect with 14 years of experience working with various large organisations
B6	17-05-2023	Business	Enterprise architect with 7 years of architecture experience at a large insurance provider
B7	19-05-2023	Technical	Director of Enterprise Architect Advisory with 16 years of experience in the banking and insurance sectors

6.1.1 Principle structure

The template utilized in this study, as presented in Table 8 draws inspiration from the well-recognized OpenGroup framework for presenting principles. Building upon the OpenGroup approach, this study extends the template with two additional elements: challenges and source.

Table 8: Design principle template

P[]	Name	[Title to identify the principle]
	Statement	[Communication of the fundamental rule]
	Rationale	[Reason for the principle]
	Implication	[Actions that should be taken]
	Challenges addressed	[Challenges that are addressed]
	Source	[Where is principle coming from]

6.2 Formulated Design principles

This section shows an overview of the overarching findings and a description of the final set of principles. It is important to note that the feedback provided by the architects through the survey has already been integrated into these principles.

6.2.1 Overarching findings

Based on the interviews conducted, several overarching findings emerged:

- Interviewees mainly come up with solutions and statements without clear principles to address the challenges. This is because there are currently no existing principles or best practices that they have utilized on this topic.
- All interviewees highlighted the absence of established best practices for addressing non-discrimination, leading to challenges in formulating principles. They drew parallels with existing principles for values like privacy.
- Interviewees emphasized the importance of regulatory bodies taking the initiative in addressing non-discrimination. They believed that regulatory involvement would prompt companies to prioritize these issues accordingly.
- The commonly proposed principles revolved around creating awareness, educating the workforce, establishing a dedicated ethics team or officer, and ensuring representative and relevant use of data.
- Not all challenges related to non-discrimination were considered adequately addressed by the proposed principles. The architects acknowledged that many of these challenges are deeply rooted in societal issues and cannot be solved solely by principles. For example, lenders intentionally biasing decisions based on subjective information to minimize their risks.
- Interviewees recognized the potential value and applicability of the proposed principles. They viewed them as tools that could address multiple challenges rather than being limited to a single issue.
- Interviewees noted that the proposed principles often overlapped, and the successful implementation of one principle may depend on the existence of another. For instance, an open culture and a dedicated team are necessary for an effective escalation path to be established and utilized. Interviewees also clearly noted that principles are applicable for multiple challenge and not only one.

6.2.2 Final set of principles

The principle of establishing an ethics team or officer within the organization is rooted in the recognition of the importance of taking proactive actions to address discriminatory issues and enforce non-discrimination principles throughout the lending process. Having this ensures that discrimination receives the necessary attention and is not overlooked or deprioritized in the organization's agenda, like how privacy is valued and safeguarded. This means that they ensure that there is dedicated attention given to discriminatory issues. Moreover, it ensures that accountability is taken. They also serve as a point of contact for reporting discrimination incidents, conducting investigations, and taking actions when necessary. Their role becomes particularly important when considering the principle of having an escalation path, as it requires a designated team to respond promptly to the reports.

P1	Name	Dedicated Ethics Team or Officer
	Statement	Establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process.
	Rationale	A dedicated team or officer helps to ensure that discrimination topics are not deprioritized or overlooked.
	Implications	• Allocate resources for the establishment and operation of the ethics teams/ officer • Establish collaboration and clear communication channels between the ethics team/ officer and other teams

		• Provide the team/ officer with authority and resources to effectively address potential discriminatory issues
	Challenges addressed	C-2, C-5, C-6, C-7
	Source(s)	B1, B3, B4, B5, B6, B7

The principle of involving diverse perspectives in a collective decision-making process is important because it helps reduce the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying solely on a single individual or viewpoint. To clarify, involving diverse perspectives helps reduce the risk of bias. Individuals naturally have their own biases, influenced by their culture and experiences, as discussed in Chapter 4. When decisions are made by a single person, these biases can unconsciously affect the outcome. However, by incorporating diverse perspectives, these biases are more likely to be recognized and counterbalanced. Moreover, conflicts of interest can also arise when decisions are made by a single individual. By involving diverse perspectives, it becomes easier to identify potential conflicts of interest and ensure that decisions are made in the best interest of the organization or customers rather than for personal gain.

P2	Name	Involve Diverse Perspectives in Collective Decision-Making
	Statement	Involve diverse perspectives in a collective decision-making process to ensure that no decision is made solely by a single individual or viewpoint.
	Rationale	The organization reduces the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying on a single individual and viewpoint.
	Implications	• Allocate resources and time for additional individuals' involvement • Clearly assign responsibility and establish accountability for decision-making outcomes • Anticipate potential delays due to the additional review step • Establish an escalation process for disagreements or conflicting viewpoints • Encourage active participation from individuals with different perspectives • Ensure that decisions are backed by data when a second decision-maker is not available • Manage communication and coordination among team members
	Challenges addressed	C-2, C-7, C-11, C-12
	Source	B1, B2, B3, B6

The principle of regularly evaluating and documenting all elements of a lending model is important for preventing discrimination by identifying and addressing potential discriminatory effects. Initially, the focus was on evaluating each individual parameter, but it has been recognized that the combination of parameters, along with the data and algorithm, can also contribute to discriminatory outcomes. Therefore, the term "model" is used as an overarching concept that includes both the algorithm, data, and parameters, as also presented in Chapter 4. Discrimination can occur unintentionally due to biased data or parameters, and regular evaluations enable organizations to detect and prevent such issues in timely manner before they escalate. Documenting these evaluations is important as it allows for easy tracing and justification of the model when required, such as during audits.

P3	Name	Evaluate the Models and their Documentation explicitly for Discriminatory Aspects
	Statement	All elements of the model (e.g., data, algorithm, parameters) used in the lending processes should be regularly evaluated and documented for potential discriminatory effects.
	Rationale	Regular evaluation of models enables organizations to identify and resolve potential issues and gain timely insight into the impact. Documentation facilitates easy tracing and justification of the model when required in the future.
	Implications	• Establish a guideline to assess the implications and alignment of models with non-discrimination objectives • Keep models up to date with newest insights on discrimination • Maintain transparency and traceability in evaluation processes • Provide training to employees involved in testing and evaluation

	Challenges addressed	C-2, C-3, C-4, C-7, C-11, C-12
	Source(s)	B1, B2, B4, B5, B6

The principle of establishing an independent entity tasked with overseeing and ensuring compliance with non-discrimination guidelines is highly beneficial. Introducing independent oversight adds an extra layer of evaluation and control, distinct from an internal ethics team. Internal teams may face conflicts of interest or be influenced by organizational pressures, potentially compromising their independence. Therefore, an external entity ensures unbiased assessment and accountability. Currently, any financial organizations tend to react to incidents of discrimination rather than proactively addressing them. This reactive approach can be attributed to the absence of proper oversight in these matters. In contrast, the field of privacy regulation has seen the establishment of independent entities like the Dutch Data Protection Authority (AP), which is responsible for enforcing and supervising compliance with the GDPR within the Netherlands. The AP possesses the authority to investigate complaints, conduct audits, and impose sanctions on organizations that violate data protection laws. Consequently, financial organizations are more inclined to proactively address privacy concerns. By implementing similar independent oversight, organizations would be more motivated to take proactive measures in preventing discrimination within their practices.

P4	Name	Independent Oversight
	Statement	Have an independent entity that is responsible for overseeing and ensuring compliance to non-discrimination guidelines.
	Rationale	The introduction of independent oversight adds an unbiased layer of evaluation and control. It enforces accountability and resolves potential conflict of interests.
	Implications	• Establish relationships with independent entities to oversee and ensure compliance • Improve and create processes to support the work of independent entities • Be open-minded and prepared to act on their findings and recommendations, which may require changes to existing processes or practices • Regular reviews may incur additional costs
	Challenges addressed	C-2, C-5, C-6, C-7
	Source(s)	B1, B6, B7

The principle of providing ongoing training and education to all employees with the aim of recognizing and addressing biases is important and is mentioned by six out of the seven architects. By educating employees and raising awareness about biases, organizations take a valuable initial step in combating discrimination. Many individuals are unaware of their own biases and lack the necessary knowledge to effectively address them. This study has revealed that most people are unaware of their biases or do not know how to respond when confronted with instances of discrimination. Providing trainings to them not only raises awareness but also equips employees with the skills and tools needed to manage and mitigate their biases. Ultimately, the organizations can foster a proactive approach to preventing discrimination, as employees become more conscious of their own biases and are better equipped to manage them effectively.

P5	Name	Create Awareness and Educate Workforce
	Statement	Continuously train and educate all employees to recognize and address biases.
	Rationale	Educating employees to raise awareness about their biases is a valuable first step, as many are unaware of their biases and lack the knowledge to address them effectively.
	Implications	• Mandate awareness training for all employees, regardless of their role or level • Evaluate and update training effectiveness and relevance regularly • Allocate resources for providing training sessions • Establish the required organizational culture and mindset • Increased time and effort required for employees to participate in mandatory trainings
	Challenges addressed	C-2, C-6, C-8, C-9

	Source(s)	B1, B3, B4, B5, B6, B7

The principle of establishing a clearly defined and well-documented escalation path for both employees and customers, coupled with a non-retaliation policy, is important. Discrimination concerns and issues often remain unresolved or inadequately addressed due to the lack of proper channels to report such matters. This is because employees and customers often do not know where and how to report instances of discrimination. In addition, even when an escalation path exists, individuals may be hesitant to come forward and report discrimination for fear of possible retaliation. By implementing a non-retaliation policy alongside a well-defined escalation path, organizations can create a safe and supportive environment that allows individuals to raise concerns without fear of negative consequences.

P6	Name	Have an Escalation Path and a Non-Retaliation Policy
	Statement	Establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy
	Rationale	Discrimination concerns and issues are not always properly raised, investigated or resolved. A clearly defined escalation path creates a safe environment for employees to report and address discrimination issues while demonstrating the organization's commitment to non-discrimination
	Implications	• Establish a structured process for raising concerns about discrimination at the appropriate level for timely resolution • Empower employees to report discriminatory issues without fear of retribution • Ensure that discriminatory issues are addressed and not ignored • Promote a culture of continuous improvement by capturing and applying lessons learned from resolving discriminatory issues
	Challenges addressed	C-4, C-5, C-9
	Source(s)	B1, B3, B4, B7

The principle of fostering an open culture where individuals are not encouraged and incentivized to report discriminatory issues and learn from mistakes is important. This is because individuals need to feel safe and encouraged to actively participate in identifying, reporting, and addressing discrimination. This promotes continuous improvement and strengthens the collective commitment to a non-discriminatory environment.

P7	Name	Culture of Openness and Learning
	Statement	Fostering an open culture where individuals are encouraged and incentivized to report discriminatory issues and learn from mistakes
	Rationale	Individuals need to feel safe and encouraged to actively participate in identifying, reporting, and addressing discrimination. This promotes continuous improvement and strengthens the collective commitment to a non-discriminatory environment and process.
	Implications	• Have processes in place that encourages a safe and open environment for individuals to speak up • Establish accessible communication channels and feedback mechanisms for reporting discrimination • Create incentives to encourage reporting and active participation in reducing discrimination • Clearly assign responsibility and establish accountability for the culture creation • Commit management and leadership to support this culture and lead by example • Update policies, processes and practices based on lessons learned • Have an escalation path in place that creates a safe environment for employees to report and address discrimination issues
	Challenges addressed	C-2, C-6, C-9
	Source(s)	B1, B2, B4, B6

The principle of traceability and understandability in all decisions regarding the lending process, both internally for employees and externally for customers and stakeholders, is important. This principle promotes transparency, facilitates independent evaluation, and promotes accountability. Currently, financial organizations frequently encounter difficulties in pinpointing the specific stages of the process that have led to disparate outcomes, primarily due to a lack of transparency and inadequate documentation for traceability. This principle also intersects with the principle of independent oversight (P4), as without well-documented, understandable, and traceable lending decisions, conducting regular audits becomes impractical for the oversight entities.

P8	Name	Traceable and Explainable Lending Process
	Statement	All decisions of the lending process should be traceable and both internally understandable by employees and externally by customers and other stakeholders.
	Rationale	Ensuring traceability, achieved through documentation and the ability to explain the tools, models, and processes utilized, is important as it enables transparency, facilitates independent evaluations, and fosters accountability.
	Implications	• Choose and design tools, models and systems to be understandable by employees • Provide training for employees especially developers to understand and explain models to non-technical users • Design systems to trace decisions back to criteria and data • Ensure an accessible and secure system for storing and managing documentation • Implement procedures on what exactly and how it should be documented • Apply regular checks to ensure correct documentation
	Challenges addressed	C-3, C-4, C-5, C-10, C-12
	Source(s)	B1, B3, B4, B6, B7

The principle of using data that is representative of the population and relevant to the lending process is important because there are currently concerns about the lack of representativeness in the data used to train models, especially when it comes to historical customer transaction data. In addition, certain data points used in determining creditworthiness, such as behavioural data, may disproportionately negatively impact individuals of a specific ethnicity, as their payment behaviour may be influenced by factors such as literacy levels. Therefore, this principle becomes essential as it stimulate organisations to thoroughly review and validate data sources and collection methods to ensure that the data used is truly representative and relevant.

P9	Name	Ensure Representative and Relevant use of Data
	Statement	Utilize data that is representative of the population and relevant to the lending processes
	Rationale	Organizations needs representative and relevant data to make well-informed and fair lending decisions.
	Implications	• Avoid relying solely on historical data • Consider collecting additional data to increase representativeness • Regularly review and validate data sources and collection methods
	Challenges addressed	C-3, C-10, C-11, C-12
	Source(s)	B1, B2, B3, B4, B5, B6

The principle of striving for undisputed decisions and actively seeking additional input when faced with disputes is important as it helps prevent potentially biased or discriminatory outcomes that could have been avoided. Individuals often make decisions solely by themselves, bringing their own biases into their judgments. Therefore, involving multiple individuals in decision-making processes, especially when uncertainties arise, becomes important.

P10	Name	Work towards Undisputed Decisions
	Statement	Strive for undisputed decisions and actively seek for additional input when faced with disputes.
	Rationale	This helps to prevent potentially biased or discriminatory outcomes that could have been avoided

	Implications	• Delay disputed decisions until there is more consensus regarding the decision at hand • Define clear process in case a decision is disputed • Include and use indicators that measure the accuracy of a system or model's output • Train algorithms and employees to identify uncertainty
	Challenges addressed	C-9
	Source(s)	B1, B4

The principle of monitoring discrimination, both direct and indirect, at every stage of the process is important as it facilitates the identification and mitigation of biases or discriminatory practices that may arise. By continuously monitoring for discrimination, financial organizations can proactively address any instances of bias that might occur throughout the process.

P11	Name	Monitor Continuously
	Statement	Discrimination, both direct and indirect, should be monitored at every stage of the process.
	Rationale	Continuous monitoring for discrimination helps to identify and mitigate biases or discriminatory practices that may emerge during the process.
	Implications	• Establish a comprehensive continuous monitoring process • Ensure monitoring mechanisms detect both direct and indirect discriminatory factors • Provide additional resources for monitoring • Report, document and communicate monitoring findings • Update policies, processes and practices based on findings
	Challenges addressed	C-2, C-3, C-4, C-7, C-8, C-9, C-11, C-12
	Source(s)	B1, B3, B4, B5

The principle of setting clear non-discrimination objectives and establishing predefined acceptance criteria that align with these objectives is important. By defining specific acceptance criteria, financial organizations can recognize non-discrimination as a goal rather than perceiving it solely as a financial burden. Currently, many organizations have started embracing sustainability as a goal rather than considering it only as a burden, which has resulted in more proactive efforts towards creating a sustainable environment. Similarly, this principle aims to achieve a similar shift in mindset, where non-discrimination becomes a recognized goal.

P12	Name	Non-Discrimination as an Objective
	Statement	Set clear non-discrimination objectives and establish predefined acceptance criteria that align with these objectives.
	Rationale	By establishing acceptance criteria, non-discrimination can be recognized as a goal rather than merely a financial burden.
	Implications	• Define clear objectives and acceptance criteria in advance, including thresholds • Utilize metrics and key performance indicators (KPIs) to track progress • Assess performance against objectives regularly • Equip decision-makers with the capabilities to effectively achieve these objectives • Continuously review and update criteria to ensure relevance
	Challenges addressed	C-2, C-6, C-9
	Source(s)	B2, B4, B5, B6, B7

6.2.2.1 Relating principles to challenges

The design principles for non-discrimination, as developed and discussed in the previous chapters, are intended to address the challenges identified in Chapter 5 and contribute to the realization of a non-discriminatory lending system. Table 9 provides a matrix that describes which principles address specific challenges. It is observed that most principles are applicable to multiple challenges, and several challenges can be addressed by multiple principles. This highlights the complexity involved in organizing and designing for non-discrimination. One challenge that has not been specifically addressed by the created design principles is C-1, which relates to lenders that can discriminate based on the information provided by borrowers. This is a typical P2P lending challenge, where lenders can make investment decisions without disclosing their criteria. As discussed in Chapter 4, some lenders justify their selectivity to reduce risk and choose borrowers they perceive as having potential and trustworthiness. The difficulty in addressing this challenge is that lenders' actions are currently beyond the control of financial organizations or regulators.

It is important to note that the implementation of these principles alone may not fully solve the issue of discrimination. The principles serve as guidance and provide a direction for organizations to follow, but their effectiveness ultimately depends on how they are implemented and integrated within each organization's specific context and practices.

Table 9: Relationships between challenges and design principles

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
C-1												
C-2	x	x	x	x	x		x				x	x
C-3			x					x	x		x	
C-4			x			x		x			x	
C-5	x			x		x		x				
C-6	x			x	x		x					x
C-7	x	x	x	x							x	
C-8					x						x	
C-9					x	x	x			x	x	x
C-10								x	x			
C-11		x	x						x		x	
C-12		x	x					x	x		x	
C-13	x			x		x					x	x

6.2.3 Ranking of principles

The architects were presented with 13 principles and asked to rank them according to their importance or relevance. The rankings provided by each architect were then averaged to determine the overall ranking of each principle. Based on the results of the interviews and the analysis of the evaluation forms, it can be observed that there is notable variability in the prioritization of the design principles among individuals. This variability highlights the subjective nature of how architects perceive and assign importance to these principles.

To illustrate this variability, a box plot (Figure 10) was created, depicting the rankings given by different architects to the principle of "monitor continuously". The rankings ranged from the top priority (1) to the least important (13), with an interquartile range of 5 to 12 and a median of 8. This wide range implies that architects differ in their assessments of the importance and relevance of the principle. Some architects consider the principle as highly relevant and deserving of a top ranking, while others may view it as less important and assign it a lower ranking. The differences in rankings could stem from variations in individual experiences, expertise, and personal values.

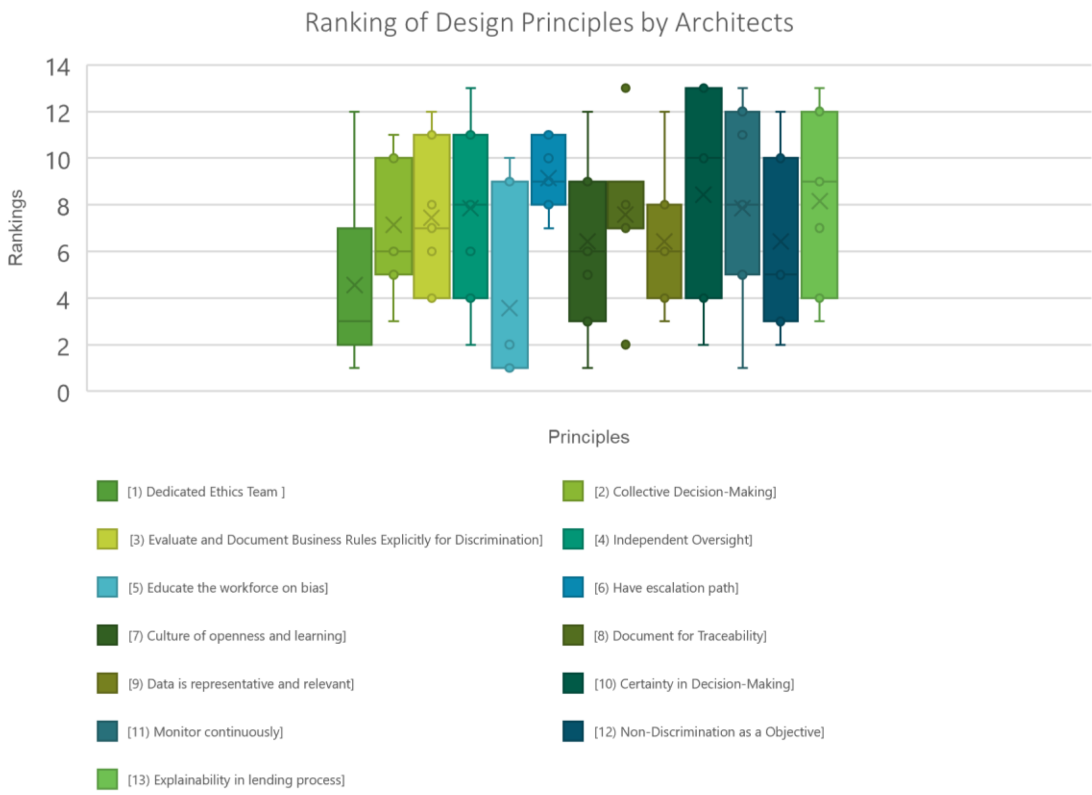


Figure 10: Boxplot of the ranking of design principles by architects

Additionally, the study did not reveal a clear distinction between technical and business architects in terms of their prioritization of the principles. Instead, it was found that architects, regardless of their technical or business background, hold unique viewpoints regarding the importance of the principles.

However, as shown in Table 10, Table 11 and Table 12. It is noteworthy that business architects rated the principle "Create Awareness and Educate Workforce" as the most important, with technical architects also ranking it highly in second place. The principle regarding having a dedicated ethics team was considered most important by technical architects. On the other hand, business architects ranked the principle of having an escalation path as the lowest priority. Technical architects ranked the principles of documenting for traceability and certainty in decision making as the lowest. This observation can be explained by the challenges technical architects face in practice regarding documentation and determining certainty.

For the remaining principles, the average rankings fell between 5 and 9, indicating a considerable variance in how architects perceived their importance. When the rankings of different architects are considered, the values tend to even out, resulting in an average that falls within this range. To clarify, if one architect rates a principle as 13 and others rate it as 1, the resulting average ranking would be 7. This example illustrates that an average ranking of 7, falling between 5 and 9, does not imply moderate importance to all architects. It emphasizes the range of individual rankings and how those differences are averaged to determine the mean. This can also be observed in the box plot figure.

However, it is important to note that due to the limited sample size of only 7 architects, it is not possible to draw conclusive findings regarding significant differences between these two groups in terms of developing and prioritizing principles. Therefore, these findings should be interpreted as an indication and solely provide guidance in prioritizing the principles. The small sample size poses a limitation, which will be discussed in the limitation section.

Table 10: Ranking of principles based on a normalized average rank from rankings provided by Business and Technical architects

Rank	Principle	Normalized average rank
1	Educate the workforce on bias	3.6
2	Dedicated Ethics Team	4.8
3	Culture of openness and learning	6.3
4	Data is representative and relevant	6.5
5	Non-Discrimination as a Objective	6.7
6	Collective Decision-Making	7.0
7	Document for Traceability	7.3
8	Evaluate and Document Business Rules Explicitly for Discrimination	7.4
9	Monitor continuously	7.9
10	Independent Oversight	8.0
11	Certainty in Decision-Making	8.2
12	Explainability in lending process	8.2
13	Have escalation path	9.2

Table 11: Ranking of principles based on the average rank provided by Business architects

Rank	Principle	Average rank	Delta to normalized average ranking
1	Educate the workforce on bias	**3.7**	0
2	Culture of openness and learning	5.3	+1
3	Document for Traceability	5.3	+4
4	Collective Decision-Making	6.0	+2
5	Certainty in Decision-Making	6.3	+6
6	Dedicated Ethics Team	6.7	-4
7	Data is representative and relevant	6.7	-3
8	Evaluate and Document Business Rules Explicitly for Discrimination	7.3	0
9	Non-Discrimination as a Objective	8.3	-4
10	Monitor continuously	8.3	-1
11	Independent Oversight	8.7	-1
12	Explainability in lending process	8.7	0
13	Have escalation path	**9.7**	0

Table 12: Ranking of principles based on the average rank provided by Technical architects

Rank	Principle	Average rank	Delta to normalized average ranking
1	Dedicated Ethics Team	**3.0**	+1
2	Educate the workforce on bias	**3.5**	-1
3	Non-Discrimination as a Objective	5.0	+2
4	Data is representative and relevant	6.3	0
5	Culture of openness and learning	7.3	-2
6	Independent Oversight	7.3	+4
7	Evaluate and Document Business Rules Explicitly for Discrimination	7.5	+1
8	Monitor continuously	7.5	+1
9	Explainability in lending process	7.8	+3

10	Collective Decision-Making	8.0	-4
11	Have escalation path	8.8	+2
12	Document for Traceability	9.3	-5
13	Certainty in Decision-Making	10.0	-2

6.3 Conclusion

This chapter focused on deriving design principles for a non-discriminatory financial lending system in collaboration with architects. Insights from interviews with seven experienced architects were used to develop principles that address the challenges identified in the previous chapter. The architects' inputs were based on their practical experience, existing guidance on safeguarding values, and potential solutions. It was found that there is a lack of specific best practices for addressing non-discrimination in financial organizations.

Key findings from the interviews emphasized the significance of regulatory involvement in tackling non-discrimination. The architects believed that regulatory bodies taking the initiative in addressing these issues would prompt companies to prioritize them accordingly. However, it was also acknowledged that the proposed principles alone cannot completely resolve all challenges related to non-discrimination, as some issues are deeply rooted in society. Nevertheless, the derived principles were considered valuable additions to existing practices and tools. It was also noted that there were overlaps among the principles, and their successful implementation relied on the presence of other complementary principles.

Through collaboration with architects and their valuable insights, a set of 12 design principles was formulated to guide the creation of a non-discriminatory financial lending system. These principles encompass various aspects such as dedicated ethics teams or officers, involving diverse perspectives in decision-making, evaluating models for discriminatory aspects, independent oversight, creating awareness and educating the workforce, establishing an escalation path and non-retaliation policy, fostering a culture of openness and learning, ensuring traceability and explainability in the lending process, using representative and relevant data, working towards undisputed decisions, continuous monitoring, and treating non-discrimination as an objective.

While multiple design principles can address most challenges, the challenge regarding lenders discriminating based on borrower information remains unaddressed as it is currently beyond the control of financial organizations and regulators.

Based on the above-described analysis, no clear distinction could be identified between technical and business architects when it comes to prioritizing these principles. It is noted that each architect has their own perspective and criteria for determining the importance of design principles. This implies that there is no agreed-upon hierarchy for these principles. Moreover, the limited sample size of 7 architects restricts the generalizability of the findings.

The next chapter will focus on evaluating the usability of these principles for improving non-discrimination in financial lending. Feedback and suggestions from another group of architects will be gathered to further refine and enhance the principles.

7 Evaluation

After the creation of design principles, it is important to evaluate their usability in improving non-discrimination in financial lending. Furthermore, gathering feedback and suggestions from other architects regarding the importance and practicality of these principles is essential to refine and sharpen them further. Therefore, the primary objective of this chapter is to address the following sRQ:

To what extent are the created design principles usable for the improvement of non-discrimination in financial lending?

This section starts by outlining the approach used for the workshop. Then it provides an overview of the overall remarks and feedback collected from both workshops. Following that, it elaborates on the suggestions received and the subsequent adjustments made to the list of principles. Finally, the framework will present the finalized list of principles. It is important to note that this study focuses on the actions that financial organizations can take to improve non-discrimination in lending. It does not address factors beyond their control. The principles and recommendations presented here are intended to guide financial organizations in addressing discrimination within their own processes and practices.

7.1 Summarized Approach

The design principles created in the previous chapter were evaluated through a combination of online and offline workshops involving a total of 4 architects (3 men and 1 woman). An overview of the participants is shown in Table 13. Another group of architects, who had not previously participated in the study, was intentionally included to prevent potential bias towards the design principles developed by the initial group. These architects were also chosen based on their relevant experience in developing principles and designing architecture specifically for financial organizations. In addition, they have a balanced combination of technical and business backgrounds.

7.1.1 Workshop Structure

Prior to the workshop, the challenges and design principles were shared with the group of architects via email. They were requested to review the principles and consider their applicability and usefulness in improving non-discrimination in financial lending. This preparatory step ensured that the discussions during the workshop could dive into the core topics and saved time.

During the interactive workshop, a PowerPoint slide was presented to clearly understand how discrimination was defined in the study and presented the identified challenges. Subsequently, each principle was presented, and the architects were instructed to focus on its usefulness in addressing discrimination and its importance and ease of applicability. Additionally, they were encouraged to improve and sharpen the formulation of the principles' name, statement, rationale, and implications when necessary. The workshop protocol can be found in Appendix J.

Moreover, the first group of architects provided feedback that the order of principles was not fully logical, and they offered suggestions for improvement. Therefore, during the evaluation session with the second group of architects, specific attention was given to determining the most logical order for these principles.

The first workshop lasted for 75 minutes, while the second workshop was slightly shorter, with a duration of 65 minutes.

Table 13: Overview of the workshop participants' expertise

	Date	ID	Expertise
Workshop 1	30-05-2023	C1	Senior manager with 10 years of experience advising financial organisations regarding their architecture
		C2	Senior advisor with 9 years of experience advising financial organisations
Workshop 2	01-06-2023	C3	Director of the enterprise architect advisory with 12 years of experience in the financial services industries
		C4	Senior advisor with 6 years of experience in various industries, including financial organisations

7.2 Feedback

This section summarizes the key remarks and feedback gathered from both workshops. It explains the suggestions received and the modifications made to the list of principles in response.

Summary of some general remarks and feedback points related to the principles:

- All participants found the principles to be highly relevant and new to them. They agree that these principles would be useful and practical for financial lending organizations.
- Concerns were raised about ensuring that discrimination becomes a priority within organizations. Participants emphasized the importance of making discrimination a central focus and ensuring that it is given the necessary attention and resources.
- Participants discussed how discrimination is not solely the responsibility of financial organizations. They provided examples, such as when a bank bases risk assessment on salary, but discrimination in salary allocation by employers indirectly leads to unequal loan outcomes. This highlighted the importance of addressing discrimination at multiple levels and involving various stakeholders.
- Participants suggested that there should be a better order or logical flow of the principles to enhance their coherence and effectiveness. They also acknowledged that there were some overlaps between the principles, but this was not seen as a major problem.
- It was highlighted that the principles are interdependent, meaning that certain principles rely on others for effective implementation. For example, an open culture is necessary for the successful implementation of an escalation path.
- The principles were considered specific to financial lending organizations but also applicable across different organizations within the financial sector. Participants appreciated that the principles were high-level and guiding, rather than prescribing specific solutions. They recognized the need for individual organizations to tailor the implementation of these principles based on their unique circumstances.

Dedicated Ethics Team or Officer

C1 highlighted that this principle can overlap with similar concepts implemented for security and privacy. C1 clarified that the primary objective of having a dedicated team in these areas is to embed security and privacy considerations throughout the entire organization. According to C1, having an advocate or an expert on such topics ensures that they are adequately taken into account in the overall software development process or organizational practices. Therefore, C1 agreed that having an ethics officer or a small team, depending on the organization's size, can contribute to improving non-discrimination efforts.

C2 also agreed with this perspective and found it concrete and actionable. Furthermore, C1 and C2 underscored the importance of this principle by emphasizing that having someone responsible for addressing non-discrimination ensures that the issue is actively flagged and addressed. They also noted that some financial organizations already have such a team or officer in place. However, ensuring non-discrimination in lending may not currently be within their scope. Therefore, it is important to ensure that existing teams also address this specific topic. Based on their suggestions, the following implication is added: ‘Ensure that the existing ethics team and officer expand their scope to address discrimination in the lending process, in addition to their current responsibilities’.

C3 and C4 agree with this principle, and C4 further discussed that it can promote proactive and reactive measures towards addressing discrimination issues. This becomes particularly important as individual borrowers or businesses may find it challenging to identify instances of discrimination, while a dedicated team with more information could be valuable in addressing this concern.

C3 added that having a dedicated team or officer is beneficial as a go-to person when ethical or discriminatory situations arise. They emphasized that currently, some financial organizations have similar positions for ethical behaviour internally, but not specifically for the lending process and its potential discriminatory outcomes. Therefore, C3 also supported this principle as it addresses the need for dedicated personnel to handle non-discrimination issues in lending.

Involve Diverse Perspectives in Collective Decision-Making

According to C1, the principle provides a clear direction that could lead to two possible solutions: (1) including diverse perspectives in the decision-making process and (2) incorporating a separate review from a discriminatory angle. C2 raised concerns about the feasibility of having an extra review step for every decision, as it may pose challenges to usability. C2 believed that the implication suggesting organizations anticipate potential delays due to the additional review step might face resistance and may not always be necessary. C1 agreed and suggested to remove that particular implication, as there are alternative solutions that do not necessarily result in delays. That implication is therefore removed.

C1 further explained that the principle is important and draws a parallel with security reviews. Organizations conduct security reviews because security issues can lead to significant risks. Similarly, if discrimination is considered important, a separate review from a discriminatory viewpoint can be included, depending on the significance of the topic and resource allocation. Therefore, it is important to identify where an additional review step is needed in the process. C1 suggested to make the principle more concrete and actionable by specifying that the additional review step should be applied only where necessary, rather than everywhere in the process. C1 further suggested a more specific approach by proposing two steps: identifying the most vulnerable areas where discrimination can occur and ensuring that these areas undergo a separate review from a diversity perspective. By incorporating these steps, the principle becomes more tangible and actionable.

Both C1 and C2 had expressed concerns regarding the practical application of this principle, noting that it relied on having a diverse organization. However, they acknowledged that achieving diversity in organizations is often challenging, making it difficult to involve diverse perspectives in the decision-making process as intended by the principle. C3 and C4 also suggested specifying that diverse perspectives are necessary for important and impactful decisions rather than every small decision, making it more practical in implementation.

Taking their feedback into consideration, the following implications are added:

- Ensure that there are diverse perspectives within the organization
- Introduce an additional review step specifically for areas that are highly vulnerable to discrimination

And the statement is changed to:

- Involve diverse perspectives in a collective decision-making for important and impactful decisions to ensure that no decision is made solely by a single individual or viewpoint

Evaluate the models and their documentation explicitly for discriminatory aspects

C1 provided an explanation of how the determination of interest rates in lending organizations involves considering the probability of customer default. Various parameters are considered in the model, focusing on those with the highest predictability. However, C1 acknowledged that these parameters can sometimes lead to discriminatory outcomes, although it is important to note that the models aim to discriminate in a legal manner. C1 agreed with the principle, as this principle directs toward the potential solution to exclude or reformulate certain parameters in the model to address discrimination. C1 suggested adding an implication to make it more practical: 'change or remove parameters from the model that are discriminatory'. C2 agreed with this perspective and noted that the original implication did not fully capture the essence of taking action based on the evaluation C2 supported this perspective and pointed out that the previous implication missed the point of taking action based on the evaluation. Therefore, the additional implication proposed by C1 was added. C1 and C2 also suggested that the principle should emphasize transparency in the model itself, not just in the evaluation process. As a result, the previous implication was revised to 'maintain transparency and traceability in model and evaluation processes.

C1 further emphasized the importance of this principle, highlighting that this area is where the key difference lies in terms of the presence of discrimination. However, C1 also noted that implementation could be challenging due to discussions that may arise. Building a model to predict the best probability of default may involve considering discriminatory parameters, and removing these parameters could result in a less accurate predictive model. This trade-off could lead to discussions and considerations regarding this principle.

C3 and C4 expressed strong agreement with this principle, particularly in the context of risk models used for both small and large loan requests. They highlighted the importance of evaluating each parameter or combination of parameters to identify any potential discriminatory outcomes. Currently, such evaluations are conducted for other factors such as profitability but not specifically for discrimination.

Independent Oversight

C1 and C2 both expressed uncertainty about the ease of implementing this principle, as typically there is a framework against which organizations are audited. Therefore, the framework itself needs to be clear first. Once clear framework is in place, it becomes relatively easier to implement the principle by conducting assessments against that framework. However, they were not aware of the existence of such a framework. Hence, an additional implication was added: 'Develop an assessment framework to assess against'. C3 and C4 agreed with this and mentioned that these should also be documented.

During the discussion, there was also some confusion regarding the interpretation of the principle as involving an external audit. However, they both agreed that an independent audit was considered important to ensure the seriousness and accountability within the organization, like financial audits. C1 and C2 explained that financial organizations usually follow a "three lines of defence" model, where the first line ensures that policies are followed based on the principle, the second line provides assurance by conducting regular checks, and the third line includes an internal audit function to verify that processes are functioning, and principles are being followed. They clarified that the internal audit function is independent. While an external audit can also be conducted, it may not necessarily add additional value in this case. To improve usability, C1 and C2 suggested adding that independent oversight can be both external and internal. C3 and C4 also agreed with this. These additions were incorporated into the statement and implications.

C3 also mentioned that the current credit risk models used by banks are evaluated by regulatory bodies like the European Banking Authority (EBA) primarily for their predictability, without explicit checks for discrimination. C3 suggested that this responsibility could be taken up by the EBA or a similar party, and this principle can contribute to driving that solution.

Create Awareness and Educate Workforce

All interviewees agreed that this principle is both easy to implement and important. However, C1 raised a remark regarding the first implication, which mandates awareness training for all employees regardless of their role or level. C1 acknowledged the relevance of the training for all employees but emphasized that the level of awareness or depth of training may vary depending on their specific functions. C1 provided an example, stating that someone working on risk models might require more comprehensive training compared to someone not directly involved in day-to-day activities. C2 agreed with this perspective. During the second evaluation workshop, they also emphasized the importance of tailoring awareness training to employees' specific roles and responsibilities to make them more concrete and actionable.

Considering their suggestion, an additional implication was added to address this point: ‘Tailor the awareness training to the specific roles and responsibilities of employees’. This modification ensures that the training provided aligns with the varying needs and requirements of employees based on their specific functions.

Have an Escalation Path and a Non-Retaliation Policy

C2 pointed out that in the Netherlands, a non-retaliation policy already exists called ‘whistle-blower’ for larger organizations. While there may be some overlap with existing whistle-blower policies, this principle can still be valuable for smaller organizations without such as policy. C1 agreed that this principle is valid, particularly because it includes internal employees and external customers. C1 suggested clarifying this inclusion in the principle's name. C2 agreed and emphasized this principle’s importance, as it also encompasses external customers. Financial organizations would be motivated to implement such a principle to safeguard their reputation and resolve complaints internally before they escalate externally. C2 added that this principle ensures closing the loop, as most principles focus on improving internal processes, but this one establishes a clear escalation path for reporting and provides a channel for customers to speak up.

They both agreed that this principle is relatively easy to implement based on their experiences with other policies, such as the whistle-blower policy. They also mentioned that this principle can be combined with existing policies. C2 agreed with this point.

During the second evaluation workshop, C3 expressed agreement with this principle, highlighting that it can prevent customers from escalating problems externally and potentially causing reputational damage to the organization. However, C3 raised a question about whether the escalation path should be within the bank that potentially discriminated or within another organization, as customers may feel uncomfortable escalating within the same bank. Nevertheless, C3 acknowledged that this principle provides a good direction for implementing an escalation path and allows for different solutions to be explored. They also suggested including ‘non-retaliation policy’ in the name to make it more explicit, as fear of retaliation can discourage individuals from reporting or discussing discriminatory issues.
Considering their feedback, the name of the principle is changed to "Have an Escalation Path for Employees and Customers, and Non-Retaliation Policy" to reflect the inclusion of both internal and external individuals and emphasize the importance of addressing concerns without fear of retaliation.

Culture of openness and learning

C2 raised a concern regarding the statement suggesting individuals should be incentivized to report issues, finding it contradictory to fostering an open environment. C2 elaborated on the concern that it might create a culture where individuals report issues solely to receive rewards, rather than foster genuine discussions and understanding. C2 did agree with the other parts of the statement of creating incentives to encourage participation and speaking up but not specifically for reporting. During the second round of interviews, both evaluators agreed with C2's concern and suggested changing the term "report" to "discuss" to ensure that discrimination is a topic that can be openly talked about and encourage open dialogue.

Furthermore, C2 found that while the implications may not be highly actionable, they are still considered adequate as they provide clear next steps. C1 and C2 also `noted that this principle overlaps with others. To address this, they suggested combining this principle with others, although they also recognized the importance of having this as a separate principle emphasizing the creation of a culture where discrimination can be openly addressed.

C1 explained that implementing this principle can be relatively easy if the organization already has an open culture. However, if the company does not have such a culture, it could pose challenges, but it is still achievable. C1 further highlighted that there have been many cases where this culture did not exist, making it extremely difficult to report discriminatory issues.

Traceable and Explainable Lending Process

C1 expressed their support for this principle, highlighting that it ensures that every step in the process is explainable, not just at the end. They emphasized the importance of considering explainability throughout the process, including in model development or other components. C2 agreed with this viewpoint, and they both believed that this principle could greatly contribute to non-discrimination efforts. When a process is traceable and explainable, individuals become more aware of their actions and its impacts. They noted that this principle works well in combination with other principles, such as the principle about evaluating the models explicitly for discriminatory aspects.

However, both C1 and C2 expressed doubts about the ease of implementing this principle, particularly when it comes to making the model itself explainable. They acknowledged the challenge of being able to trace back which parameters contributed to specific outcomes within the model. Nonetheless, they agreed that if it is possible to achieve explainability, it is undoubtedly useful and important. C2 further emphasized that if a model is not understood by anyone, other principles such as fostering an open and learning culture become ineffective, as there would be nothing to discuss. Additionally, they noted the importance of making the lending process explainable for external oversight.

C4 regarded this principle as one of the most important, as it promotes awareness of potential problems in the lending process from the beginning. Understanding the process allows for consideration of consequences and facilitates better decision-making. C3 agreed with this viewpoint.

No specific changes or remarks were mentioned regarding this principle, indicating agreement with its content and importance.

Ensure Representative and Relevant use of Data

C1 initially suggested changing the wording from "relevant use of data" to "exclude non-relevant data" as they believed that excluding non-relevant data during model training is more logical. However, during the second workshop, C3 and C4 disagreed with this change. They argued that determining in advance which data is non-relevant, especially during the development of the model, is not feasible, and excluding data based on assumptions of irrelevance could lead to loss of potential insights or later discoveries of relevance. They emphasized the need to use data appropriately rather than excluding it without proper evaluation. Therefore, the original wording of "relevant use of data" was retained. Moreover, during the interviews for sub question 1, A3, who is an expert in bias in algorithm and data, explicitly mentioned that removing data from training models is not the solution as it can skew the data and make it less accurate. It is also challenging to identify in advance which data may lead to discriminatory outcomes, especially due to the presence of proxies.

C1 and C2 acknowledged the importance of this principle. However, they also recognized that compared to some other principles, such as the principle about having an ethics team or officer, implementing this principle may be more challenging. C1 highlighted the difficulty in determining which parameters to exclude, particularly when considering variables like postal codes that may indirectly discriminate against certain groups. These grey areas can lead to extensive discussions and considerations. During the second workshop, C3 further supported this point and explained that in some cases, certain areas may have higher accident rates, making the use of postal codes relevant for accurate risk calculations. However, there is also a dilemma as postal codes can be indirectly linked to certain characteristics, raising concerns of potential discrimination. However, they all found it implementable and important to include in the list of principles.

C3 also recognized the problem that the data used is not always representative as it uses historical data from 10 to 15 years ago. Moreover, making the data set smaller will make the model less reliable. Hence, C3 also admitted that the current data used for risk models are always a little behind regarding the representation of the society or customer group at that moment and do not know how to make it more representative, but this principle can lead to solutions to address this issue.

Work towards Undisputed Decisions

C2 expressed confusion regarding the usability of the principle of working towards undisputed decisions in improving non-discrimination. They did not understand how using indicators to measure the accuracy of a system or model's output could lead to working towards discriminatory decisions. C2 believed that undisputed decisions could be useful for any issue and did not specifically contribute to reducing discrimination. C1 shared a similar perspective, stating that while it could potentially be usable, they were unsure if it held the same level of importance as other principles. C1 also highlighted potential conflicts with other goals, as waiting for undisputed decisions could lead to significant delays in the decision-making process, impacting the interests of customers. Additionally, determining the threshold for certainty and when a decision becomes undisputed poses challenges.

During the second evaluation workshops, C3 also expressed that it is not always feasible to have undisputed decisions, as it would mean that decisions could never be made if there is any dispute. They found this principle less straightforward and clear compared to others.

Considering the feedback from both groups and the lack of agreement on the usefulness and clarity of this principle, it was decided to remove it from the set of design principles. This decision was also influenced by the fact that more architects disagreed with the principle than agreed, with only B1 and B4 expressing clear support for it.

Monitor Continuously

C1 initially mentioned that this principle could be redundant as it relates to the principle regarding independent oversight, as continuous monitoring is part of the oversight function. C2 agreed with this perspective and questioned the need for a separate principle, suggesting that combining it with other principles might be more appropriate.

However, during the second workshop, C3 and C4 strongly opposed the idea of removing or combining this principle with others. They believed that it was essential to have a separate principle explicitly emphasizing the need for continuous monitoring, rather than treating it as an implicit component of other principles. They argued that the responsibility of the ethics team or independent oversight team in ensuring ongoing monitoring can result as a solution derived from this principle.

Considering the input from the four architects (B1, B3, B4, and B5) who proposed this principle and the strong support from the second evaluation group to include it as a separate principle, it was decided to retain this principle, despite the initial suggestion to remove or combine it.

Non-Discrimination as an Objective

C1 and C2 did not have remarks about this principle, but they suggested changing the order so that it becomes the first principle. They discussed the importance of utilizing metrics and KPIs to track progress, as it implicitly emphasizes the need for monitoring, which is related to the principle of continuous monitoring.

However, C3 expressed concerns about the measurability of this principle. They found it challenging to define specific metrics for tracking progress. In response, C4 provided an insightful perspective, noting that even if the exact measurement methods are currently unclear, the principle's importance lies in giving a clear direction. C4 drew a parallel with the concept of making sustainability measurable, which may have seemed challenging in the past but has seen progress over time. This highlights the potential for finding ways to measure progress when effort is dedicated to it, even if the solutions are not immediately apparent.

Overall, the feedback from C1, C2, C3, and C4 indicates general agreement on the importance of this principle, with some discussions on the order of principles and the measurability aspect.

7.3 Final set of principles

After conducting the evaluation of the design principles and incorporating the feedback received, the final design principle framework is presented below. Appendix K shows the adjustments.

Based on the feedback received during the first workshop and the acknowledged need for a more practical and logical arrangement of the principles, the architects from the second workshop were assigned the task of categorizing the principles. The objective was to create a framework that would bring about a clearer classification of the principles and enhance their practical application. Subsequently, the categorized principles and their order were thoroughly discussed with the architects from the first group. This iterative process continued until a consensus was reached on the categorization. The outcome of this collaborative effort is the development of the following principles framework:

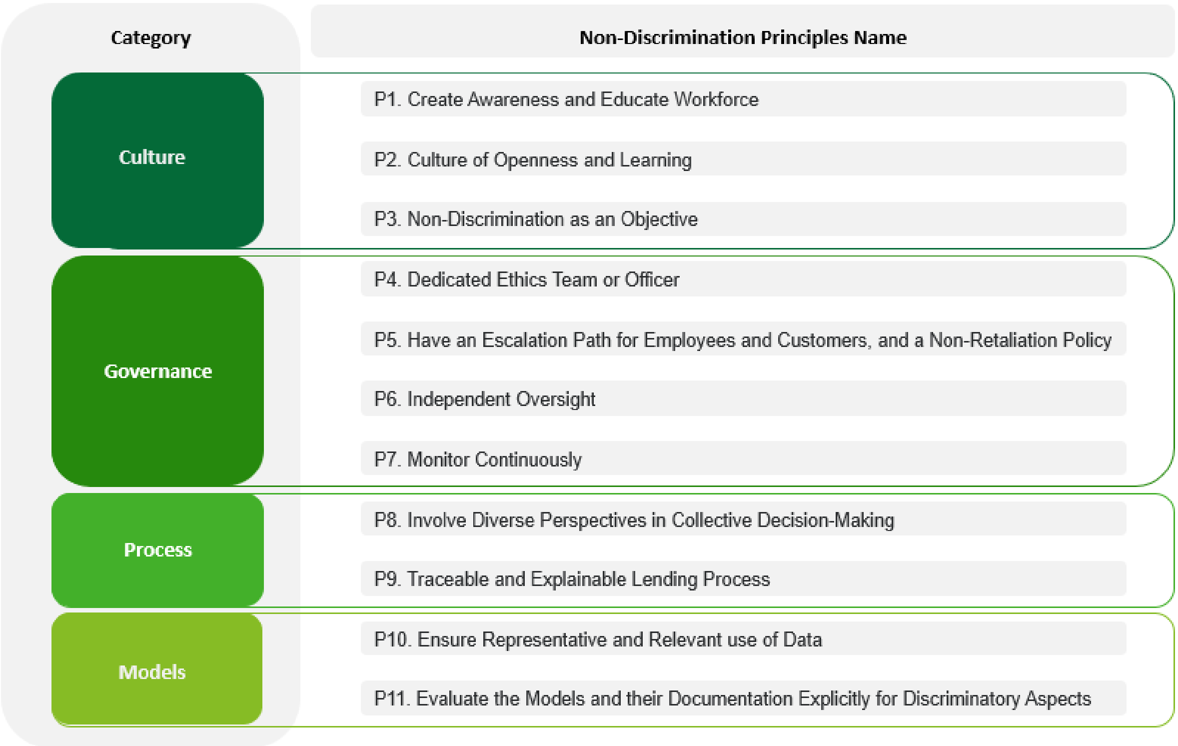


Figure 11: Framework of non-discrimination principles

Culture

P1	Name	Create Awareness and Educate the Workforce
	Statement	Continuously train and educate all employees to recognize and address biases.
	Rationale	Educating employees to raise awareness about their biases is a valuable first step, as many are unaware of their biases and lack the knowledge to address them effectively.
	Implications	• Mandate awareness training for all employees, regardless of their role or level • Tailor the trainings to the specific roles and responsibilities of employees. • Evaluate and update training effectiveness and relevance regularly • Allocate resources for providing training sessions • Establish the required organizational culture and mindset • Increased time and effort required for employees to participate in mandatory trainings

P2	Name	Culture of Openness and Learning
	Statement	Fostering an open culture where individuals are encouraged and incentivized to discuss discriminatory issues and learn from mistakes
	Rationale	Individuals need to feel safe and encouraged to actively participate in identifying, discussing, and addressing discrimination. This promotes continuous improvement and strengthens the collective commitment to a non-discriminatory environment and process.
	Implications	• Have processes in place that encourages a safe and open environment for individuals to speak up • Establish accessible communication channels and feedback mechanisms for discussing discrimination • Create incentives to encourage discussions and active participation in reducing discrimination • Clearly assign responsibility and establish accountability for the culture creation • Commit management and leadership to support this culture and lead by example • Update policies, processes and practices based on lessons learned • Have an escalation path in place that creates a safe environment for employees to discuss and address discrimination issues

P3	Name	Non-Discrimination as an Objective
	Statement	Set clear non-discrimination objectives and establish predefined acceptance criteria that align with these objectives.
	Rationale	By establishing acceptance criteria, non-discrimination can be recognized as a goal rather than merely a financial burden.
	Implications	• Define clear objectives and acceptance criteria in advance, including thresholds • Utilize metrics and KPIs to track progress • Assess performance against objectives regularly • Equip decision-makers with the capabilities to effectively achieve these objectives • Continuously review and update criteria to ensure relevance

Governance

P4	Name	Dedicated Ethics Team or Officer
	Statement	Establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process.
	Rationale	A dedicated team or officer helps to ensure that discrimination topics are not deprioritized or overlooked.
	Implications	• Allocate resources for the establishment and operation of the ethics teams/ officer • Establish collaboration and clear communication channels between the ethics team/ officer and other teams • Provide the team/ officer with authority and resources to effectively address potential discriminatory issues • Ensure that the existing ethics team and officer expand their scope to address discrimination in the lending process, in addition to their current responsibilities

P5	Name	Have an Escalation Path for Employees and Customers, and a Non-retaliation Policy
	Statement	Establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy
	Rationale	Discrimination concerns and issues are not always properly raised, investigated or resolved. A clearly defined escalation path creates a safe environment for employees to report and address discrimination issues while demonstrating the organization's commitment to non-discrimination
	Implications	• Implement a non-retaliation which can also be achieved by leveraging existing non-retaliation policies if they are already in place. • Establish a structured process for raising concerns about discrimination at the appropriate level for timely resolution • Empower employees to report discriminatory issues without fear of retribution • Ensure that discriminatory issues are addressed and not ignored • Promote a culture of continuous improvement by capturing and applying lessons learned from resolving discriminatory issues.

P6	Name	Independent Oversight
	Statement	Have an external or internal independent entity that is responsible for overseeing and ensuring compliance to non-discrimination guidelines.
	Rationale	The introduction of independent oversight adds an unbiased layer of evaluation and control. It enforces accountability and resolves potential conflict of interests.
	Implications	• Establish relationships with external or internal independent entity to oversee and ensure compliance • Improve and create processes to support the work of independent entities • Be open-minded and prepared to act on their findings and recommendations, which may require changes to existing processes or practices

		• Regular reviews may incur additional costs • Develop an assessment framework to assess against

P7	Name	Monitor Continuously
	Statement	Discrimination, both direct and indirect, should be monitored at every stage of the process.
	Rationale	Continuous monitoring for discrimination helps to identify and mitigate biases or discriminatory practices that may emerge during the process.
	Implications	• Establish a comprehensive continuous monitoring process • Ensure monitoring mechanisms detect both direct and indirect discriminatory factors • Provide additional resources for monitoring • Report, document and communicate monitoring findings • Update policies, processes and practices based on findings

Process

P8	Name	Involve Diverse Perspectives in Collective Decision-Making
	Statement	Involve diverse perspectives in a collective decision-making for important and impactful decisions to ensure that no decision is made solely by a single individual or viewpoint.
	Rationale	The organization reduces the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying on a single individual and viewpoint.
	Implications	• Allocate resources and time for additional individuals' involvement • Clearly assign responsibility and establish accountability for decision-making outcomes • Establish an escalation process for disagreements or conflicting viewpoints • Encourage active participation from individuals with different perspectives • Ensure that decisions are backed by data when a second decision-maker is not available • Manage communication and coordination among team members • Ensure that there are diverse perspectives within the organization • Introduce an additional review step specifically for areas that are highly vulnerable to discrimination

P9	Name	Traceable and Explainable Lending Process
	Statement	All decisions of the lending process should be traceable and both internally understandable by employees and externally by customers and other stakeholders.
	Rationale	Ensuring traceability, achieved through documentation and the ability to explain the tools, models, and processes utilized, is important as it enables transparency, facilitates independent evaluations, and fosters accountability.
	Implications	• Choose and design tools, models and systems to be understandable by employees • Provide training for employees especially developers to understand and explain models to non-technical users • Design systems to trace decisions back to criteria and data • Ensure an accessible and secure system for storing and managing documentation • Implement procedures on what exactly and how it should be documented • Apply regular checks to ensure correct documentation

Models

P10	Name	**Ensure Representative and Relevant use of Data**
	Statement	Utilize data that is representative of the population and relevant to the lending processes
	Rationale	Organizations needs representative and relevant data to make well-informed and fair lending decisions.
	Implications	• Avoid relying solely on historical data • Consider collecting additional data to increase representativeness • Regularly review and validate data sources and collection methods

P11	Name	**Evaluate the Models and their Documentation explicitly for Discriminatory Aspects**
	Statement	All elements of the model (e.g., data, algorithm, parameters) used in the lending processes should be regularly evaluated and documented for potential discriminatory effects.
	Rationale	Regular evaluation of models enables organizations to identify and resolve potential issues and gain timely insight into the impact. Documentation facilitates easy tracing and justification of the model when required in the future.
	Implications	• Establish a guideline to assess the implications and alignment of models with non-discrimination objectives • Keep models up to date with newest insights on discrimination • Maintain transparency and traceability in model and evaluation processes • Provide training to employees involved in testing and evaluation • Change or remove parameters from the model that are discriminatory

7.4 Conclusion

This chapter focused on evaluating the principles established in the previous chapter in terms of their applicability and usefulness in mitigating discrimination in financial lending. The evaluation was conducted through discussions in two workshops. Overall, the principles were considered useful and applicable for financial organizations. They provided flexibility for organizations to adapt the principles to their specific circumstances. However, one principle, which emphasized on working towards undisputed decisions, was found to be less relevant in directly addressing discrimination. The feasibility of always reaching undisputed decisions was questioned, as it could potentially impede decision-making processes. It was also recognized that discrimination is a complex issue that extends beyond the scope of financial organizations alone. External factors, such as income inequality, can indirectly influence loan outcomes and contribute to discrimination.

Based on the feedback received in the workshops, efforts were made to improve the structure and consistency of the principles. The principles were rearranged and grouped into four categories: culture, governance, process, and models. This reorganization aims to improve the understanding and applicability of the principles.

8 Discussion and Conclusion

This chapter concludes this study by discussing the findings to answer the mRQ. Section 8.1 discusses the findings. Section 8.2 answers the mRQ. Section 8.3 explores the scientific contribution, while Section 8.4 delves into the societal and managerial contribution of the study. Section 8.5 discusses the study's limitations, while Section 8.6 offers recommendations for future research. Section 8.7 reflects on the research process. Lastly, Section 8.8 explains the connection between the study and the CoSEM program.

8.1 Discussion on findings

Throughout the research, multiple new findings and insights were discovered on this topic which enhances the overall understanding of the topic. The accumulation of these new information gradually reshaped the initial perception of the researcher on the topic. The study underlined the role of context, highlighting how discrimination differs in each context based on regulations, stakeholders' needs, and other factors. The research also discovered numerous tensions, such as the constraints placed by the GDPR. This regulation prevents companies from collecting and processing certain types of sensitive data. This restriction creates a tension that complicates the process of determining whether certain outcomes are indirectly related to these sensitive data, potentially leading to discriminatory outcomes. Another tension emerges from the subjective judgments made by lenders about discrimination. While some view certain practices as discriminatory, others may deem them necessary to mitigate risks. This brings to light the wide variance in perceptions of discrimination, depending on one's stance and perspective. These tensions and other findings will be discussed in more details in this section.

8.1.1 The existence of P2P lending

The study started with the idea that P2P lending might help to reduce discrimination in financial lending. It seemed to offer an alternative mean for people who could not get a loan from a traditional bank. This idea is also supported by existing literature. However, the findings in this study indicated that while P2P lending improves loan accessibility, it does not necessarily decrease discrimination.

The findings indicated that lenders often base their decisions on factors beyond borrower's control, such as ethnicity, age, and physical appearance. A big concern is that these lenders do not need to provide an explanation for their decisions, potentially resulting in discriminatory outcomes. It was even found that P2P lending might inadvertently enable discrimination, as lending companies can sidestep responsibility for unequal outcomes by placing the lending decision and responsibilities in the hands of individual lenders.

It was also discovered that most of the research in the field of P2P lending primarily focuses on platforms operating in the USA and China, while European platforms are underrepresented. This lack of literature can be attributed to the relatively limited presence of P2P lending platforms in Europe, especially in the Netherlands. This study found that there are several reasons contributing to the limited presence of P2P lending platforms in the Netherlands. The financial system in the country is highly regulated and emphasizes on financial well-being, with maximum interest rates and strict regulations that make it less financially attractive for P2P companies to operate there. Additionally, the Dutch financial lending context differs from other countries, as no credit agencies can provide P2P lending platforms with credit scores to build risk models. Consequently, assessing their customers' risks becomes challenging for P2P lending companies, requiring them to develop their own credit scoring systems.

However, developing their own credit scoring systems presents another challenge for P2P lending companies as they often lack historical data on their customers, which traditional financial institutions use to build their risk models. Banks and other financial institutions collect these customer data as part of their regular business operations and they cannot share this data with other organizations without consent or legal justifications, such as GDPR regulations. As a result, P2P lending companies face limitations in obtaining the necessary historical data. However, relying solely on historical data can result in worse loan terms for borrowers who do not fit within the population represented in the historical data. This situation is currently observed with traditional financial institutions. This presents an opportunity for P2P lending companies to have a positive impact by developing their own risk models that are different from traditional ones, including a new group of previously excluded individuals. Additionally, some P2P lending platforms now use non-traditional data sources like social media profiles and online behaviour to enhance their credit scoring models. While these sources offer potential benefits, it is important to address ethical implications and prevent other potential discrimination.

In addition, there were discussions on the potential benefits of P2P lending for the overall financial well-being of society and particularly on individuals who do not qualify for traditional banks. These banks use rigorous risk assessments to assess borrowers' ability to repay loans in which they consider various factors. This approach helps mitigate the risk of borrowers taking on loans they cannot afford, ensuring responsible lending. Turning to P2P lending allows individuals who do not meet banks' strict criteria to create additional risks. P2P lending platforms can increase the likelihood of borrowers taking on loans they cannot afford, leading to potential financial instability. This raises the question of whether P2P lending is the optimal solution or if traditional financial institutions should focus on addressing discrimination issues and developing more inclusive lending practices.

8.1.1.1 Context dependent

This study found that P2P lending is context dependent and its usefulness is different for each country. In Indonesia, where many small micro-enterprises, like street sellers, are excluded from accessing loans through traditional banks, P2P lending emerges as a valuable alternative. Despite being capable of repaying loans, these individuals are often considered risky by banks due to their informal status. Therefore, P2P lending provides them with an opportunity to access financing. In contrast, the Netherlands possesses a well-developed and regulated financial system with established banks, making credit more accessible through traditional channels. The study indicates that while most individuals in the Netherlands can generally obtain bank loans, there may be instances where individuals struggle to secure sufficient lending, particularly for mortgages. This situation raises questions about alternative options like P2P lending, which could potentially fill the gaps left by traditional banks. However, there is a dilemma associated with relying on P2P lending as a solution in such cases. The concern arises from the potential negative impact on borrowers' financial well-being. This poses risks not only to borrowers, but also to the stability of the P2P lending platform and lenders involved.

This shows the need to carefully consider both advantages and disadvantages of P2P lending in different contexts. It is important to find a balance between enabling financial inclusion and ensuring responsible lending practices to protect interests of both borrowers and lenders.

8.1.2 The complexity of discrimination

This study developed a conceptual framework that showed the complexity of discrimination. This framework itself serves as a finding as it brings together various perspectives on the issue, shedding light on the interconnected nature of discrimination. Initially, the focus was on examining discrimination within the architecture of P2P lending. However, as the research progressed, it became clear that discrimination is not just an architectural problem and cannot be solved by financial organisations alone. Discrimination manifests itself in various domains, including regulatory and human aspects. These domains and their associated factors are interconnected, meaning that changes in one factor, such as altering a regulation, can have an impact on the motivation of organizations and the development of processes and models within them. Moreover, conversations with various individuals showed that everyone has their own view of discrimination and its root causes. Some interviewees attributed discrimination to organizational practices, while others blamed regulators. This lack of consensus shows the need to look at the problem of discrimination from a broader perspective that transcends the boundaries of one domain.

8.1.3 The different interpretations of discrimination

This study found that (P2P) lenders are primarily driven to invest when the financial returns from utilizing platforms outweigh the additional risk involved and surpass other investment opportunities with similar levels of risk. Lenders tend to focus predominantly on their return on investment and risk mitigation, often neglecting potential discriminatory practices. Furthermore, the perception of discrimination can be subjective, as individuals may interpret certain practices as discriminatory while others may view them as reasonable measures to mitigate risks. This wide variation in the perception of discrimination among different stakeholders can be attributed to their unique perspectives shaped by their roles and experiences. This suggests the need for a standard or guideline that goes beyond individual viewpoints and define what constitutes discrimination in the context of lending.

8.1.4 Legal vs non-legal normative agencies and regulations

This study found that the contrasting influences of legal and non-legal normative agencies impact how organizations address and prioritize discrimination issues. While discrimination is illegal, the judgement of organizations such as the Netherlands Institute for Human Rights are not legally binding. In contrast, privacy concerns are guided by the legally binding GDPR, which is enforced by the Dutch Data Protection Authority. The GDPR establishes clear and enforceable penalties for non-compliance, enforcing companies to comply strictly with privacy regulations. This disparity in enforcement mechanisms highlights the differential attention given to privacy and discrimination. It underscores the power of legal and non-legal normative agencies and regulations in shaping corporate behaviour, as organizations tend to prioritize issues that have direct legal consequences.

It would therefore be recommended to delve into the examination of how discrimination issues are handled in other countries. While this study focused on the Netherlands, it would be valuable to explore jurisdictions where discrimination enforcement measures or authorities are in place and assess their effectiveness, drawing valuable lessons from them.

8.2 Conclusion

This section summarizes the conclusions of each sRQ of this study and answers the mRQ.

The study started with identifying the factors contributing to financial discrimination in the lending system. To make the concept of discrimination more applicable in the context of financial lending, the study defined specific dimensions: differential treatment, differential outcome, unequal access, and lack of transparency. These dimensions provide a framework for understanding how discrimination can manifest in the lending system and how they can interact with one another.

Furthermore, the study identified and categorized various factors that can contribute to one or more dimensions of discrimination into six domains: regulatory, human interaction, user experience, business, data, and technical. These domains help identify the specific areas in which discrimination can arise within the lending context and how different factors may overlap across multiple domains.

Having gained an understanding of the dimensions, factors, and domains of discrimination, the study proceeded to identify the challenges encountered in realizing a non-discriminatory financial lending system in the Netherlands. This was accomplished through a combination of SLR and interviews, which identified 13 challenges. The aim of recognizing these challenges was to acknowledge the obstacles and difficulties associated with achieving a non-discriminatory system.

To answer the mRQ:

What are design principles for a non-discriminatory financial lending system?

The study developed design principles that could address these challenges. These principles were also evaluated in two workshops involving other group of architects. Based on the feedback received, efforts were made to improve the structure and consistency of the principles, resulting in a final set of 11 principles. To enhance clarity, the principles were rearranged and grouped into four categories: culture, governance, process, and models.

The design principles for realizing a non-discriminatory financial lending system are described as follows:

Culture:

- Principle 1 is aimed to continuously train and educate all employees to recognize and address biases. Educating employees about their biases raises awareness and equips them with the knowledge to effectively address biases that may exist within the lending system.
- Principle 2 is aimed to foster an open culture where individuals are encouraged and incentivized to discuss discriminatory issues and learn from mistakes. This promotes an environment where individuals feel safe to actively participate in identifying, discussing, and addressing discrimination, leading to continuous improvement and a collective commitment to a non-discriminatory environment and process.
- Principle 3 is aimed to set clear non-discrimination objectives and establish predefined acceptance criteria that align with these objectives. By establishing acceptance criteria, non-discrimination is recognized as a goal rather than merely a financial burden, ensuring that decisions are made in accordance with these objectives.

Governance:

- Principle 4 is aimed to establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process. Having a dedicated team or officer ensures that discrimination topics are not deprioritized or overlooked.
- Principle 5 is aimed to establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy. This creates a safe environment for reporting and addressing discrimination concerns, demonstrating the organization's commitment to non-discrimination.
- Principle 6 is aimed to have an external or internal independent entity responsible for overseeing and ensuring compliance with non-discrimination guidelines. Independent oversight adds an unbiased layer of evaluation and control, enforces accountability, and resolves potential conflicts of interest.
- Principle 7 is aimed to monitor discrimination, both direct and indirect, at every stage of the lending process. Continuous monitoring helps identify and mitigate biases or discriminatory practices that may emerge during the process, ensuring that fair and equitable practices are upheld.

Process:

- Principle 8 is aimed to involve diverse perspectives in collective decision-making for important and impactful decisions. By including diverse perspectives, the organization reduces the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying on a single individual or viewpoint.

- Principle 9 is aimed to ensure that all decisions made in the lending process are traceable and understandable both internally by employees and externally by customers and other stakeholders. Ensuring traceability, which can be achieved through documentation and the ability to explain the tools and models utilized, enables transparency, facilitates independent evaluations, and fosters accountability.

Models:

- Principle 10 is aimed to utilize data that is representative of the population and relevant to the lending processes. Organizations require representative and relevant data to make well-informed and fair lending decisions.
- Principle 11 is aimed to regularly evaluate and document all elements of the model (e.g., data, algorithm, parameters) used in the lending processes for potential discriminatory effects. Regular evaluation of models enables organizations to identify and resolve potential issues, while documentation facilitates easy tracing and justification of the model when required in the future.

It is important to note that there are overlaps among the principles, and their successful implementation relied on the presence of other complementary principles.

The study acknowledges that while the identified design principles can contribute to reducing discrimination in the financial lending system, they do not fully resolve the problem. This is because many discriminatory issues are deeply rooted in society and are caused by multiple factors. However, the principles derived from this study are valuable additions to existing practices and tools, providing a framework for organizations to improve their lending practices. In addition, the involvement of the regulatory authorities is essential to address discrimination in the lending system. Without regulatory incentives or oversight, organizations are not motivated to proactively implement such principles or other initiatives aimed at promoting non-discriminatory lending practices.

8.3 Scientific contribution

The proposed set of principles for addressing discrimination within financial organizations is considered the most important contribution of this study to the existing literature. While most ethical literature critically examines discrimination and identifies underlying challenges, it often lacks detailed design guidance for effectively addressing these issues. This study serves as a good starting point in filling this gap by providing a concrete set of design principles that financial organizations can utilize to address discrimination. These principles offer practical guidance and actionable steps for organizations to implement to address discriminatory lending practices. It is recommended that these principles be employed at the beginning of the design process or during a transformation of an architecture to ensure that non-discriminatory practices are integrated from the beginning. Moreover, this study contributes to the literature by using the PBD science methodology. By using PBD, the study demonstrates the suitability of this methodology for deriving principles for socio-technical systems which aligns with the findings of Bharosa and Janssen (2015).

This study also contributes to the existing research by taking a socio-technical lens to examine discrimination. It sheds light on the complexity of discrimination within the financial lending system and highlights the involvement of organizations outside the financial sector, such as employers, in contributing to discriminatory outcomes. Demonstrating the interplay and influence between different factors emphasizes the importance of studying discrimination holistically rather than isolating the factors and domains. This study could encourage other researchers to see the importance of a broader perspective in addressing discrimination-related issues.

This study also discovered many factors and their relationships that need to be further explored in future research. These findings provide opportunities for in-depth research into various factors of discrimination. In addition, future research could delve into ways to deal with the challenges identified in this study. These insights from the study form a basis for future researchers to delve deeper into the complexity of discrimination.

8.4 Societal and managerial contribution

During interviews with professionals from the financial industries and advisory sectors, it became evident that there is currently a lack of established design principles for non-discrimination in financial lending. This observation highlights the societal importance of having such principles in place, as they ensure that non-discrimination is integrated in the organization's design right from the outset. Adhering to these principles enables companies to comply with existing discrimination regulations and prepares them for future regulatory developments in this field.

These principles can also serve as a source of inspiration for policymakers to develop universal principles or other policies to address discrimination issues more broadly. As discussed in this study, in the absence of regulatory bodies addressing discrimination and establishing a set of mandatory principles, like the GDPR for privacy, companies will lack the motivation to take proactive measures in combating discrimination. However, as mentioned in the knowledge gap section, there are policies that indirectly touch upon discrimination. Section 8.4.1 examines the principles developed in this study against three significant policy documents in the AI ethics field that partially address discrimination. The goal is to understand the similarities with existing principles and evaluate how they can enrich the current guidelines.

Moreover, the design principles developed to address discrimination in lending have the potential to be applied in other areas, such as governmental financial allowances, extending their impact beyond the financial lending system. A recent article from NOS on June 23, 2023, highlighted concerns about the fairness of fraud controls related to student grants administered by DUO, a Dutch governmental organization responsible for distributing these grants and loans. An investigation revealed a disproportionate number of accusations against students with a migrant background, raising concerns about the selection process and potential discrimination. Criticism has been directed at DUO's algorithm-based system for detecting fraud, which lacks a scientific basis and relies on the biased experiences of employees, potentially leading to the targeting of specific groups and resulting in false accusations. The government is currently exploring ways to address this issue, these principles can be generalized to this context and serve as a source of inspiration for deriving solutions to address the discrimination problem.

8.4.1 Reflection on existing policy

This section reflects on the main findings of this study in relation to three significant policy documents regarding AI ethics: the Ethics Guidelines for Trustworthy AI, the EGE Group Report on AI Robotics and Autonomous Systems, and the IEEE Global Initiative on Ethics of AI and Intelligent Systems. The aim is to identify similarities and differences, focusing on what these principles contribute to the field of AI ethics.

Both the developed principles and the mentioned policy documents share key themes such as transparency, explainability, accountability, and non-discrimination. These principles are widely recognized as fundamental to ethical AI. An additional insight from this research, however, is the link between explainability, transparency, and accountability in creating non-discriminatory systems. While these concepts are discussed individually in the policy documents, this study presents them as interconnected factors essential to addressing discrimination.

There are notable differences between the principles derived from this research and those in the broader policy documents. The developed principles are distinct in their focus, specificity, and industry relevance. Existing guidelines are broad and can be considered unclear, which could lead to varied interpretations. For instance, in the Ethics Guidelines for Trustworthy AI, one of the statements is that "unfair bias must be avoided". However, it might be unclear what constitutes "unfair bias" or how to address it in practical terms. Additionally, their wide-ranging focus on multiple ethical issues may spread the attention thin over non-discrimination. In contrast, the principles developed in this study are more practical and dedicated to addressing discrimination. They provide more clear direction to ensure non-discrimination, such as the suggestion of a dedicated ethics team or officer and the establishment of a clear escalation path. These concrete steps can help organizations directly address and prevent discriminatory practices. Additionally, each principle comes with its defined set of implications, further enhancing their applicability and impact. This is currently missing with the policy documents.

Another advantage of these principles is their industry-specific focus. They have been developed with the lending system in mind, which allows them to address unique challenges within this context, a level of specificity that current general AI ethics guidelines do not offer. Moreover, these principles can also be generalized to other industries as discussed before.

However, it is important to clarify that the existing policy documents are valuable as they guide organizations in dealing with AI ethics. The principles from this study serve to enrich these broader principles and can be viewed as one of the derivatives of them. Also, these principles are not limited to AI; they can be applied in various other areas beyond AI.

8.5 Limitations

Due to time constraints, not all aspects of this study were researched in detail. This section outlines the limitations of this study, pinpointing areas where the findings may necessitate cautious interpretation and offering guidance for future research.

Limitations of the developed framework

The conceptual framework encompassing various factors of discrimination was presented based on the expertise and experiences of the interviewed individuals. The connections between these factors were included when mentioned by at least one interviewee. However, it is important to note that despite the knowledge and experience of each interviewee in its respective field, it is not yet clear or proven whether each arrow's impact is statistically significant. Additionally, this research did not delve into determining the precise influence of each factor on one another. Another limitation to consider is that the framework itself has not been validated, which could affect its reliability and applicability.

Unexplored interactions

This study's focus was solely on examining whether the identified design principles can improve non-discrimination, without exploring any potential interactions between these principles and other design principles. Consequently, it is uncertain how these principles will translate into practical applications. It is important to acknowledge that these design principles can potentially have broader implications for other values, such as profitability, as noted by the interviewees. The interviewees noted that implementing these principles will require trade-offs and considerations to ensure alignment with other business objectives.

Architecture framework selection

Another limitation of this study is that it employed a specific layered framework to guide targeted and structured data gathering by identifying suitable representatives for each layer. This approach facilitated the collection of data that was more focused and directly relevant to the research objectives. However, it is important to acknowledge that utilizing alternative frameworks, such as incorporating additional layers not considered in this study, may yield different outcomes. For instance, introducing a security layer could uncover additional factors or challenges that may manifest within that specific layer. Therefore, the findings and conclusions of this study should be interpreted within the scope of the chosen framework, recognizing that alternative frameworks may provide different perspectives and insights.

Limited Dutch P2P lending perspectives

Another limitation of this study is that it did not include interviews with P2P lending companies specifically in the Netherlands. While the presence of P2P lending platforms in the country is relatively limited, efforts were made to approach one platform that identifies itself as P2P, although it leans more towards crowdfunding. Unfortunately, they did not respond to the research inquiry.

Limitation of the chosen sample

Another limitation of the study is that all the interviewed architects belonged to the same advisory company. While this allowed for a broad industry perspective based on their experience and exposure to various clients, it raises concerns about the potential lack of diversity in the perspectives gathered. The architects could have similar biases that could influence the information collected and perspectives obtained, potentially leading to principles that are more specific to their company rather than widely applicable across the industry. However, it is important to note that despite belonging to the same company, the architects in this research presented different principles and held distinct perspectives on what principles could be useful in addressing the problem. This suggests that individual variation and diverse viewpoints were still present within the sample.

Insufficient sample for architect comparison

Another limitation of this study is that it examined the potential differences in prioritizing design principles between business and technical architects. Although the findings did not reveal substantial differences and the main variations observed were based on individual differences. Moreover, due to the limited sample size of only 7 architects, it was not possible to conclude whether there is a significant difference between these two groups in terms of developing and prioritizing principles.

8.6 Recommendations for future research

This section presents the recommendations for future research, drawing on both the findings and limitations of this study.

The role of policymakers in promoting non-discriminatory practices

The developed principles in this research can serve as practical guidelines for organizations aiming to foster non-discriminatory practices. However, a consistent concern raised by various experts and practitioners was the challenge of motivating or ensuring that companies will implement these principles. This challenge is primarily because organizations often have limited capacity and must comply with numerous regulatory demands within tight timeframes. For example, existing strict regulations related to privacy, KYC, and ESG have led many financial organizations to prioritize these issues, leaving others, such as reducing bias in risk models, neglected. This highlights the strong influence of regulations on an organization's task prioritization.

Organizations typically react to regulatory demands rather than proactively addressing issues. When facing regulatory sanctions, such as fines, organizations become more focused and motivated to take accountability or prioritize the task at hand. Therefore, policymakers should recognize the need to address discrimination specifically and implement measures that enforce organizations to prioritize non-discriminatory practices. While the study showed that discrimination is a complex issue, extending beyond just the regulatory domain, all interviewees agreed that without stricter regulations, organizations might not take the initiative on their own. This pattern is like how organizations responded to issues of privacy and sustainability in the past. Future studies can explore and assess the steps and measures policymakers can take to ensure organizations actively work towards reducing discrimination.

The conceptual framework

The developed conceptual framework, illustrating the complexity of discrimination, is composed of various domains, factors and relationships. As noted in the limitations section, not every aspect was thoroughly researched, primarily due to time constraints. The impact of each connection between the factors within the framework has not yet been determined or validated. Future research can be conducted to delve into each of these connections to better understand how the factors influence one another and to determine their significance. Additionally, it would be useful for future research to focus on identifying any potential missed domains, factors, or relationships that were not initially captured.

Moreover, conducting this study with a different group of interviewees would provide an opportunity to assess whether a similar framework emerges, providing insights into the reliability of the findings. Furthermore, it would be valuable to explore whether similar domains, factors, and relationships exist in other countries. This could be done by conducting comparative studies across different countries to shed light on the generalizability and transferability of the framework.

The existence of P2P lending

This study provides an initial exploration into the absence of P2P lending companies in the Netherlands and the various factors contributing to this phenomenon. The findings highlighted the complexity of the situation, with strict regulations, the existing financial system, and the lack of necessary data all playing a role. However, given the benefits that P2P lending companies have brought to other countries, it would be useful for future researchers and policymakers to investigate further. Specifically, the potential benefits of P2P lending incentivization in the Netherlands can be investigated. Required changes to implement P2P lending can be a possible follow-up RQ if benefits are deemed sufficient. This could include a detailed analysis of the regulatory environment, an assessment of potential benefits and risks, and a consideration of how P2P lending could fit in the broader financial landscape of the Netherlands.

Validating the principles

The principles in this research were primarily driven and evaluated by the insights of a select group of senior advisors. Future research could include architects working directly in financial organizations or seek input from a broader range of stakeholders within the industry to derive principles to address the potential bias and lack of diversity in perspectives. Another direction for future research is to validate the derived principles by applying them to real or hypothetical scenarios. By testing the principles in various scenarios, researchers can further enhance the validity and applicability of the findings.

Furthermore, while some formulations of the principles may be predominantly focused on financial contexts, it is important to recognize that they can still offer valuable insights in other domains. By rephrasing and adapting these principles, they could be effectively applied to diverse scenarios, including governmental financial allowances or other settings beyond the financial lending system.

8.7 Reflection on research process

The journey through this study has been filled with numerous learnings and insights, not only about the topic but also about the research process itself. The researcher strongly believe DSR was the suitable method to use for this topic. This is because the understanding of discrimination and the development of design principles were more complex than expected due to the complexity of discrimination. It required many adjustments and improvements along the way. DSR is a method that allowed me to keep improving my discrimination framework and principles through feedback from the environment and insights from literature. This led to a more comprehensive view on the issue.

However, in hindsight, a couple improvements could be made to the research process. One key lesson from this journey is the importance of a clear initial definition of the research problem. The excitement of starting the research led to an early dive into data gathering through interviews and desk research. While this gave vast amount of information, it did make for a challenging course later due to a lack of clear focus. A sharper definition of the research problem from the beginning would have helped streamline the process and ensure a more efficient use of time.

The desk research led to many insightful findings due to the many scientific articles included in this research. The temptation to read extensively on the subject was driven by the researcher's interest for the topic and the fear of missing out on important information. However, this experience highlighted the importance of prioritizing quality over quantity. The approach adopted led to repeated information, and recognizing this, the researcher has learned that a well-reasoned and explained methodology for selecting relevant and high-quality literature can suffice. Moving forward, adopting a more selective strategy would indeed be more beneficial, without compromising the research's integrity.

The process of collecting data through interviews taught the researcher another important lesson about balance. While having a lot of data can provide diverse perspectives, it also created a vast amount of data that demanded weeks for analysis. There were indeed nights where sleep was sacrificed to manage the overflowing information. In the future, a more selective approach in choosing interview participants would be adopted to ensure a manageable analysis process, while also ensuring sufficient rest.

Finally, one of the most crucial lessons the researcher grasped was to embrace the complexity of research. As an engineer, the researcher was trying to solve every problem and question that arose. However, a professor helped the researcher to understand that discovering new challenges and acknowledging the complexity itself is already a valuable finding and contribution to research. This realization has allowed the researcher to appreciate that research goes beyond simply finding solutions; it is a journey of exploration and understanding.

8.8 Link with CoSEM

This thesis project showed and addressed the complexities of discrimination in the financial lending system through a socio-technical perspective. It went beyond understanding the technical factors of discrimination, but also considered regulatory and human interactions. This shows that the study required a combination of technical, institutional, economic, and social knowledge to understand the complexity of discrimination.

In addition, the study successfully navigated the diverse and sometimes even conflicting perspectives (e.g., academic, legal, industrial, etc.) on discrimination and visually illustrated its complexity through a conceptual framework. Based on this understanding, the study developed design principles to address the problem. This demonstrates the typical problem-solving design approach of the CoSEM program.

9 References

Abbassi, P., & Iyer, R. (2021). Asymmetric Information at the Platform Level: The Case of Peer-To-Peer Lending. *The Palgrave Handbook of Technological Finance*, 407–423. https://doi.org/10.1007/978-3-030-65117-6_17

Abdelmaksoud, N. S. A. (2022). Universal Design Approach to a Smart, Sustainable, Safe, and Accessible City for All. *WIT Transactions on Ecology and the Environment*, *260*, 259–270. https://doi.org/10.2495/SC220211

Adams, W., Einav, L., & Levin, J. (2009). Liquidity Constraints and Imperfect Information in Subprime Lending. *American Economic Review*, *99*(1), 49–84. https://doi.org/10.1257/AER.99.1.49

Alshammari, M., & Simpson, A. (2017). Towards a principled approach for engineering privacy by design. *Privacy Technologies and Policy: 5th Annual Privacy Forum*, *10518 LNCS*, 161–177. https://doi.org/10.1007/978-3-319-67280-9_9

Atkins, R., Cook, L., & Seamans, R. (2022). Discrimination in lending? Evidence from the Paycheck Protection Program. *Small Business Economics*, *58*(2), 843–865. https://doi.org/10.1007/S11187-021-00533-1/TABLES/11

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, L., & Tiburtius, P. (2011). Online peer-to-peer lending—A literature review. *Journal of Internet Banking and Commerce*, *16*(2). https://www.researchgate.net/profile/Burkhardt-Funk/publication/236735575_Online_Peer-to-Peer_Lending--A_Literature/

Bennett, C. J. (2018). The European General Data Protection Regulation: An instrument for the globalization of privacy standards? *Information Polity*, *23*(2), 239–246. https://doi.org/10.3233/IP-180002

Bharosa, N., & Janssen, M. (2015). Principle-Based Design: A Methodology and Principles for Capitalizing Design Experiences for Information Quality Assurance. *Journal of Homeland Security and Emergency Management*, *12*(3), 469–496. https://doi.org/10.1515/JHSEM-2014-0073/MACHINEREADABLECITATION/RIS

Borgesius, F. J. Z., & Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, *24*(10), 1572–1593. https://doi.org/10.1080/13642987.2020.1743976

Buit, M. E. (2020). Consumer peer-to-peer lending and the promise of enhancing access to credit: Lessons from the Netherlands. *Discrimination, Vulnerable Consumers and Financial Inclusion*, 166–184. https://doi.org/10.4324/9781003055075-10

Burgstahler, S. E., & Cory, R. C. (2010). Universal Design in Higher Education: From Principles to Practice. In *Harvard Education Press*. Harvard Education Press. 8 Story Street First Floor, Cambridge, MA 02138. Tel: 888-437-1437; Tel: 617-495-3432; Fax: 978-348-1233; e-mail: hepg@harvard.edu; Web site: http://hepg.org/hep-home/home.

Cardinal, A., Gonzales, L., & Rose, E. J. (2020). Language as Participation: Multilingual User Experience Design. *Proceedings of the 38th ACM International Conference on Design of Communication*, 1–7. https://doi.org/10.1145/3380851.3416763

Cavoukian, A., Taylor, S., & Abrams, M. E. (2010). Privacy by Design: essential for organizational accountability and strong business practices. *Identity in the Information Society 2010 3:2*, *3*(2), 405–413. https://doi.org/10.1007/S12394-010-0053-Z

Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., & Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 3721–3726. https://doi.org/10.1073/PNAS.0710069105/SUPPL_FILE/IMAGE1974.GIF

Chakraborty, P., & Sarkar, A. (2017). Context driven approach for enterprise architecture framework. *Computer Information Systems and Industrial Management: 16th IFIP TC8 International Conference*, *10244 LNCS*, 277–289. https://doi.org/10.1007/978-3-319-59105-6_24/FIGURES/5

Chander, A. (2016). The Racist Algorithm. *Michigan Law Review*, *115*. https://heinonline.org/HOL/Page?handle=hein.journals/mlr115&id=1081&div=&collection=

Chen, X., Huang, B., & Ye, D. (2020). Gender gap in peer-to-peer lending: Evidence from China. *Journal of Banking & Finance*, *112*, 105633. https://doi.org/10.1016/j.jbankfin.2019.105633

Cherednychenko, O. O. (2017). The EU Charter of Fundamental Rights and Consumer Credit: Towards Responsible Lending? In *European Contract Law and the Charter of Fundamental Rights* (pp. 139–166). https://papers.ssrn.com/abstract=2971194

Clark, N., Gill, J., O'Dea, E., Ramamani, R., Ravi, R., & Saha, S. (2023). *Horizon architecture: The hidden superpower for adapting to change and winning in the market*. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/strategy/us-horizon-architecture-the-hidden-superpower-for-adapting-to-change-and-winning-in-the-market-part-one.pdf

Clegg, C. W. (2000). Sociotechnical principles for system design. *Applied Ergonomics*, *31*(5), 463–477. https://doi.org/10.1016/S0003-6870(00)00009-0

Corten, P. A. (2018). *Implementation of Blockchain Powered Smart Contracts in Governmental Services (MSc thesis)*. TU Delft.

Cumming, T. M., & Rose, M. C. (2022). Exploring universal design for learning as an accessibility tool in higher education: a review of the current literature. *Australian Educational Researcher*, *49*(5), 1025–1043. https://doi.org/10.1007/S13384-021-00471-7/FIGURES/1

Daniel, A., Sreekumar, K., & Minu, K. K. (2014). A Survey on Direct and Indirect Discrimination Prevention in Data Mining. *International Journal of Computer Applications*, *108*(14). https://research.ijcaonline.org/volume108/number14/pxc3900376.pdf

Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design Science Research. *Design Science Research*, 67–102. https://doi.org/10.1007/978-3-319-07374-3_4

Duarte, J., Siegel, S., & Young, L. (2012). Trust and Credit: The Role of Appearance in Peer-to-peer Lending. *The Review of Financial Studies*, *25*(8), 2455–2484. https://doi.org/10.1093/RFS/HHS071

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, *47*(1), 54–70. https://doi.org/10.1080/00036846.2014.962222

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c741934bd883bac1b1446b8e4886ae03295c001a

European Commission. (2023, March 22). *Non-discrimination*. https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/equality/non-discrimination_en

Fairlie, R., Robb, A., & Robinson, D. T. (2022). Black and White: Access to Capital Among Minority-Owned Start-ups. *Management Science*, *68*(4), 2377–2400. https://doi.org/10.1287/MNSC.2021.3998

Ferati, M., Mripa, N., & Bunjaku, R. (2016). Accessibility of MOOCs for blind people in developing non-English speaking countries. *Advances in Intelligent Systems and Computing*, *500*, 519–528. https://doi.org/10.1007/978-3-319-41962-6_46/TABLES/2

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance Decision Making: A Field Study of Prosocial Lending. *Journal of Marketing Research*, *48*(SPL), S130–S137. https://doi.org/10.1509/JMKR.48.SPL.S130

Garz, S., Giné, X., Karlan, D., Mazer, R., Sanford, C., & Zinman, J. (2021). Consumer Protection for Financial Inclusion in Low- and Middle-Income Countries: Bridging Regulator and Academic Perspectives. *Annual Review of Financial Economics*, *13*, 219–246. https://doi.org/10.1146/ANNUREV-FINANCIAL-071020-012008

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending. *Ournal of Management Information Systems*, *34*(2), 401–424. https://doi.org/10.1080/07421222.2017.1334472

Gimpel, H., Heger, S., & Wöhl, M. (2022). Sustainable behavior in motion: designing mobile eco-driving feedback information systems. *Information Technology and Management*, *23*(4), 299–314. https://doi.org/10.1007/S10799-021-00352-6/TABLES/3

Gjøsaeter, T., Radianti, J., & Chen, W. (2021). Towards Situational Disability-aware Universally Designed Information Support Systems for Enhanced Situational Awareness. *ISCRAM 2020 Conference Proceedings–17th International Conference on Information Systems for Crisis Response and Management. International Association for Information Systems for Crisis Response and Management*. https://oda.oslomet.no/oda-xmlui/bitstream/handle/10642/9506/ISCRAM2020-2295_TerjeGjosaeter_etal2021.pdf

Goecks, L. S., de Souza, M., Librelato, T. P., & Trento, L. R. (2021). Design Science Research in practice: review of applications in Industrial Engineering. *Gestão & Produção*, *28*(4), 2021. https://doi.org/10.1590/1806-9649-2021V28E5811

Gonzalez, L. (2020). Blockchain, herding and trust in peer-to-peer lending. *Managerial Finance*, *46*(6), 815–831. https://doi.org/10.1108/MF-09-2018-0423/FULL/PDF

Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, *2*, 44–58. https://doi.org/10.1016/J.JBEF.2014.04.002

Greiner, M. E., & Wang, H. (2010). Building consumer-to-consumer trust in E-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce*, *15*(2), 105–136. https://doi.org/10.2753/JEC1086-4415150204

Harkness, S. K. (2016). Discrimination in Lending Markets. *Social Psychology Quarterly*, *79*(1), 81–93. https://doi.org/10.1177/0190272515623459

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*(5), 837–852. https://doi.org/10.1037/0022-3514.78.5.837

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell Me a Good Story and I May Lend you Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *Journal of Marketing Research*, *48*(Special Issue), S138–S149. https://doi.org/10.1509/JMKR.48.SPL.S138

Iversen, T. N., Larsen, L., & Solem, P. E. (2012). A conceptual analysis of Ageism. *Nordic Psychology*, *61*(3), 4–22. https://doi.org/10.1027/1901-2276.61.3.4

Iyer, R., Khwaja, A. I., Luttmer, E. F. P., & Shue, K. (2012). Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.1570115

Iyer, R., Khwaja, A. I., Luttmer, E. F. P., & Shue, K. (2016). Screening Peers Softly: Inferring the Quality of Small Borrowers. *Management Science*, *62*(6), 1554–1577. https://doi.org/10.1287/MNSC.2015.2181

Karakaxas, A., Karakostas, B., & Zografos, V. (2000). A business object oriented layered enterprise architecture. *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, *IEEE 2000*, 807–810. https://doi.org/10.1109/DEXA.2000.875118

Klafft, M. (2008). Online Peer-to-Peer Lending: A Lenders' Perspective. *Proceedings of the 2008 International Conference on E-Learning, e-Business, Enterprise Information Systems, and e-Government, EEE 2008*, 371–375. https://doi.org/10.2139/SSRN.1352352

Korstjens, I., & Moser, A. (2018). European Journal of General Practice Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, *297*(3), 1083–1094. https://doi.org/10.1016/J.EJOR.2021.06.023

Kruglanski, A. W., & Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, *13*(1), 1–44. https://doi.org/10.1002/EJSP.2420130102

Kumar, I. E., Hines, K. E., & Dickerson, J. P. (2022). Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation. *Proceedings of the 2022*

AAAI/ACM Conference on AI, Ethics, and Society, 357–368. https://doi.org/10.1145/3514094.3534154

Lange, M., Mendling, J., & Recker, J. (2012). Measuring the realization of benefits from Enterprise Architecture Management. *Journal of Enterprise Architecture*. https://eprints.qut.edu.au/50620/16/50620.pdf

Lee, M., & Floridi, L. (2021). Algorithmic Fairness in Mortgage Lending: from Absolute Conditions to Relational Trade-offs. *Minds and Machines*, *31*(1), 165–191. https://doi.org/10.1007/S11023-020-09529-4/TABLES/2

Lenz, R. (2016). Peer-to-Peer Lending: Opportunities and Risks. *European Journal of Risk Regulation*, *7*(4), 688–700. https://doi.org/10.1017/S1867299X00010126

Lincoln, Y., & Guba, E. (1985). *Naturalistic Inquiry*. Sage.

Luo, C., Xiong, H., Zhou, W., Guo, Y., & Deng, G. (2011). Enhancing investment decisions in P2P lending: An investor composition perspective. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 292–300. https://doi.org/10.1145/2020408.2020458

Lytvynenko, T., Tkachenko, I., Ivasenko, V., & Lvovska, T. (2020). Application of the Universal Design Principles in the Improvement of Street and Urban Road Environment. *Lecture Notes in Civil Engineering*, *73*, 143–150. https://doi.org/10.1007/978-3-030-42939-3_16/FIGURES/4

Mach, T., Carter, C., & Slattery, C. R. (2014). *Peer-to-peer lending to small businesses*.

Mack, K., Bragg, D., Morris, M. R., Bos, M. W., Albi, I., & Monroy-Hernández, A. (2020). Social App Accessibility for Deaf Signers. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 31. https://doi.org/10.1145/3415196

Manor, S., & Herscovici, A. (2021). Digital ageism: A new kind of discrimination. *Human Behavior and Emerging Technologies*, *3*(5), 1084–1093. https://doi.org/10.1002/HBE2.299

Matheus, R., Janssen, M., & Janowski, T. (2021). Design principles for creating digital transparency in government. *Government Information Quarterly*, *38*(1), 101550. https://doi.org/10.1016/J.GIQ.2020.101550

Meuser, M., & Nagel, U. (2009). The Expert Interview and Changes in Knowledge Production. *Interviewing Experts*, 17–42. https://doi.org/10.1057/9780230244276_2

Montanha, I. R., Matiello, J. P., Roqueiro, N., Da Rosa, E., Nicolazzi, L. C., Ogliari, A., & De Souza Vieira, R. (2008). Guidelines for vehicle development based on principles of universal design. *SAE Technical Papers*. https://doi.org/10.4271/2008-36-0257

Morse, A., & Pence, K. (2021a). Technological Innovation and Discrimination in Household Finance. *The Palgrave Handbook of Technological Finance*, 783–808. https://doi.org/10.1007/978-3-030-65117-6_28

Morse, A., & Pence, K. (2021b). *Technological Innovation and Discrimination in Household Finance*. 783–808. https://doi.org/10.3386/W26739

Mosca, E. I., & Capolongo, S. (2020). Universal Design-Based Framework to Assess Usability and Inclusion of Buildings. *ICCSA 2020: 20th International Conference*, *12253 LNCS*, 316–331. https://doi.org/10.1007/978-3-030-58814-4_22/TABLES/2

Nasef, E. M. M., & Bakar, N. A. B. (2020). Enterprise Architecture “As-Is” Analysis for Competitive Advantage A Case Study of the Internet Service Provider Company. *International Journal of Advanced Computer Science and Applications*, *11*(7), 1–6. https://thesai.org/Downloads/Volume11No7/Paper_14-Enterprise_Architecture_As_Is_Analysis.pdf

Noble, S. U. (2018). *Algorithms of Oppression*. New York University Press. https://doi.org/10.18574/nyu/9781479833641.001.0001

Orhan, M. (2019). Women Business Owners in France: The Issue of Financing Discrimination. *Journal of Small Business Management*, *39*(1), 95–102. https://doi.org/10.1111/0447-2778.00009

Pittman, C. N., & Heiselt, A. K. (2014). Increasing Accessibility: Using Universal Design Principles to Address Disability Impairments in the Online Learning Environment. *Online Journal of Distance Learning Administration*, *17*(3).

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture?: Evidence of Discrimination from Prosper.com. *Journal of Human Resources*, *46*(1), 53–92. https://doi.org/10.1353/jhr.2011.0025

Prystav, F. (2016). Personal information in peer-to-peer loan applications: Is less more? *Journal of Behavioral and Experimental Finance*, *9*, 6–19. https://doi.org/10.1016/J.JBEF.2015.11.005

Qiu, J., Lin, Z., & Luo, B. (2012). Effects of borrower-defined conditions in the online peer-to-peer lending market. *Lecture Notes in Business Information Processing*, *108 LNBIP*, 167–179. https://doi.org/10.1007/978-3-642-29873-8_16/COVER

Rau, P. R. (2018). Law, Trust, and the Development of Crowdfunding. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2989056

Ravina, E. (2019). Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.1101647

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Seidel, S., Chandra Kruse, L., Székely, N., Gau, M., & Stieger, D. (2018). Design principles for sensemaking support systems in environmental sustainability transformations. *European Journal of Information Systems*, *27*(2), 221–247. https://doi.org/10.1057/s41303-017-0039-0

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*(2), 253–263. https://doi.org/10.1016/S0377-2217(01)00113-8

Soydaş Çakır, H., & Tosun, A. (2022). Universal design approach to analysis of physical environment for users with multiple sclerosis. *Journal of Accessibility and Design for All*, *12*(2), 284–319. https://doi.org/10.17411/JACCES.V12I2.339

Sukmaningsih, D. W. (2018). A Model for Lender-Borrower Trust in Peer-To-Peer Lending. *ComTech: Computer, Mathematics and Engineering Applications*, *9*(1), 15. https://doi.org/10.21512/COMTECH.V9I1.4287

Sulastri, R., & Janssen, M. (2022). The elements of the Peer-to-peer (P2P) lending system: A Systematic Literature Review. *15th International Conference on Theory and Practice of Electronic Governance*, *1*, 424–431. https://doi.org/10.1145/3560107.3560172

Sulastri, R., & Janssen, M. (2023). *Measuring financial inclusiveness in credit financing*.

United Nations. (2018). *The 2030 Agenda and the Sustainable Development Goals*. https://sdgs.un.org/2030agenda

Van den Hoven, J., Vermaas, P. E., & Van de Poel, I. (2015). *Handbook of Ethics, Values, and Technological Design*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6970-0

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*(1), 1–18. https://doi.org/10.1186/S12874-018-0594-7/TABLES/3

Wachter, S., Mittelstadt, B., & Russell, C. (2020). Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law. *West Virginia Law Review*, *123*. https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=6331&context=wvlr

Wang, H., Chen, K., Zhu, W., & Song, Z. (2015). A process model on P2P lending. *Financial Innovation*, *1*(1), 1–8. https://doi.org/10.1186/S40854-015-0002-9/FIGURES/8

Wang, M., Vogel, D., & Ran, W. (2011). Creating a performance-oriented e-learning environment: A design science approach. *Information & Management*, *48*(7), 260–269. https://doi.org/10.1016/J.IM.2011.06.003

Wijeratne, D., Dennehy, D., Quinlivan, S., Buckley, L. A., Keighron, C., & Flynn, S. (2022). Learning Without Limits: Identifying the Barriers and Enablers to Equality, Diversity, and Inclusion in IS Education. *Journal of Information Systems Education*, *33*(1), 61–74. https://aisel.aisnet.org/jise/vol33/iss1/8

Wu, B., Liu, Z., Gu, Q., & Tsai, F. S. (2023). Underdog mentality, identity discrimination and access to peer-to-peer lending market: Exploring effects of digital authentication. *Journal of International*

Financial Markets, Institutions and Money, *83*, 101714. https://doi.org/10.1016/J.INTFIN.2022.101714

Yang, Q., & Lee, Y. C. (2016). Critical factors of the lending intention of online P2P: Moderating role of perceived benefit. *ACM International Conference Proceeding Series*, *17-19-August-2016*. https://doi.org/10.1145/2971603.2971618

Yiangkamolsing, C., Bohez, E. L. J., & Bueren, I. (2010). Universal design (UD) principles for flexible packaging and corresponding minimal customer requirement set. *Packaging Technology and Science*, *23*(5), 283–300. https://doi.org/10.1002/PTS.900

Yum, H., Lee, B., & Chae, M. (2012). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Electronic Commerce Research and Applications*, *11*(5), 469–483. https://doi.org/10.1016/j.elerap.2012.05.003

Appendix A

This appendix provides a more detailed explanation of the legal definition of discrimination, as discussed in the Theoretical Background chapter.

Legal definition of discrimination

The EU has several directives against discrimination that must be implemented by member states, including the Netherlands. These directives include Directive 2000/43/EC against race and ethnic origin discrimination, Directive 2000/78/EC against workplace discrimination based on religion or belief, disability, age, or sexual orientation, Directive 2006/54/EC for equal treatment of men and women in employment and occupation, Directive 2004/113/EC for equal treatment of men and women in access to and supply of goods and services, and Directive Proposal (COM(2008)462) against discrimination beyond the workplace based on age, disability, sexual orientation, and religion or belief (European Commission, 2023).

These non-discrimination directives have been incorporated into Dutch law. The Netherlands primarily implemented these directives through Article 1 of the Dutch Constitution, which states that all people in the Netherlands are equally treated in equal circumstances, and discrimination on the grounds of religion, belief, political opinion, race, gender, or any other grounds is not permitted. While Article 1 does not provide detailed rules on protected characteristics, EU discrimination directives are transposed into Dutch law through more comprehensive and specialized legislation, such as the Dutch General Equal Treatment Act (Wet AWGB, 2023), which bans discrimination on various grounds, including belief, gender, civil status, nationality, political opinion, race, religion, and sexual orientation. The Equal Treatment (Men and Women) Act further elaborates on distinctions between men and women. Additionally, Dutch law addresses disability through the Equal Treatment of Disabled and Chronically Ill People Act and age through the Equal Treatment in Employment (Age Discrimination) Act. Although not explicitly targeting financial services, these acts apply across all sectors, including the financial industry. Moreover, the Financial Supervision Act, which oversees the majority of the Netherlands' financial sector, mandates that providers must grant access to the financial system based on objective criteria and transparent, non-discriminatory rules. This analysis results in a list of 11 distinct variables derived from regulatory documents. The overview of all the variables is presented in Table 1.

Table A1: The variables considered in both EU directives and Dutch law

	EU Directive					Dutch Law				
Variables	Directive 2000/43/EC	Directive 2000/78/EC	Directive 2006/54/EC	Directive 2004/113/EC	Directive Proposal (COM(2008)462)	Article 1 of the Dutch Constitution	The General Equal Treatment Act	The Equal Treatment in Employment (Age Discrimination) Act	The Equal Treatment of Disabled and Chronically Ill People Act	The Equal Treatment (Men and Women) Act
Belief		x			x	x	x			
Gender			x	x		x	x			x
Civil status	x						x			
Nationality	x						x			
Political opinion						x	x			
Race	x					x	x			
Religion		x			x	x	x			

Sexual orientation		x			x		x			
Age		x			x			x		
Disabled		x			x				x	
Chronologically ill					x				x	

Appendix B

This Appendix shows some additional insights on the data collection process for sRQ 1.

Desk research

Table B1 shows an overview of the used articles related to the borrowers and lenders perspective. Most of them (Yes) are derived from the query and the others (No) are derived from snowballing.

Table B1: Overview of all used articles in the borrower and lender perspective research

Authors	Title	From query?
(Prystav, 2016)	Personal information in peer-to-peer loan applications: Is less more?	Yes
(Yang & Lee, 2016)	Critical factors of the lending intention of online P2P: Moderating role of perceived benefit	Yes
(Emekter et al., 2015)	Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending	Yes
(Herzenstein et al., 2011)	Tell Me a Good Story and I May Lend you Money: The Role of Narratives in Peer-to-Peer Lending Decisions	Yes
(Galak et al., 2011)	Microfinance Decision Making: A Field Study of Prosocial Lending	No
(Ravina, 2019)	Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets	No
(Sukmaningsih, 2018)	A Model for Lender-Borrower Trust in Peer-To-Peer Lending	No
(Ge et al., 2017)	Borrower's default and self-disclosure of social media information in P2P lending	Yes
(Lenz, 2016)	Peer-to-Peer Lending: Opportunities and Risks	No
(Iyer et al., 2016)	Screening Peers Softly: Inferring the Quality of Small Borrowers	No
(Gonzalez, 2020)	When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans	No
(Duarte et al., 2012)	Trust and Credit: The Role of Appearance in Peer-to-peer Lending	Yes
(Qiu et al., 2012)	Effects of borrower-defined conditions in the online peer-to-peer lending market	Yes
(Pope & Sydnor, 2011)	What's in a picture?: Evidence of discrimination from Prosper.com	Yes
(Klafft, 2008)	Online Peer-to-Peer Lending: A Lenders' Perspective	No
(Cesarini et al., 2008)	Heritability of cooperative behavior in the trust game	No
(Mach et al. 2014)	Peer-to-peer lending to small businesses Board of Governors of the Federal Reserve System	No

Interviews

This section describes the potential risks of interviews and the expert identification process.

Potential risks

Semi-structured interviews were chosen as the research method for addressing specific RQ, namely sRQ 1, sRQ 2, and sRQ 4. This method is chosen as it allows for flexibility and adaptability in capturing the perspectives and expertise of the interviewees while ensuring the research objectives are met. The interviews were designed to be open-ended, allowing for in-depth exploration of the topics, while also guided by an interview protocol that ensured the key areas of interest were covered.

In addition to the benefits of flexibility and standardized questions, there are also some disadvantages associated with using semi-structured interviews as a research method. The interviewees may not fully explore all aspects of their experiences and best practices, and the findings from the qualitative content analysis will also be limited by the sample size of the interviews and the subjectivity of the analysis process. To reduce these limitations, it is recommended that sufficient interviews are conducted to ensure that a wide range of experiences and perspectives are captured. Vasileiou et al. (2018) found that a minimum sample size of 12 is recommended for qualitative research to achieve data saturation. In this study, a total of 13 interviews were conducted, which can be considered an appropriate sample size for the scope of the research. In case not enough experts could be found, alternative methods such as online surveys were considered. If these methods were not feasible, the findings from the literature review would have been complemented with expert opinions sourced from grey literature. However, this was not necessary for this study.

To address the limitation of the smaller sample size in developing the design principles, the researchers implemented two strategies. First, the seven architects who were initially interviewed also evaluated each other's principles, providing cross-validation and refinement. Second, four additional architects who were not part of the initial study were invited to evaluate the design principles through an evaluation workshop. These steps were taken to involve a larger number of participants and reduce potential biases.

Expert identification

It is important to identify the individuals who are considered experts and will be interviewed for this study. According to Meuser & Nagel (2009), an expert is someone who possesses institutionalized authority in their field of action and has exclusive knowledge that is not accessible to others in that field. They also highlighted that the decision on who is appropriate for an expert interview is made by the researcher, taking into consideration their research objectives and the recognition of the individual as an expert in their own field. Shanteau et al. (2002) found that there are various methods to determine whether someone is an expert. These include years of experience, certification, social recognition, consistency, consensus, discrimination ability, behavioural traits, knowledge tests, and creation of experts. However, each of these methods has one or more significant limitations, and there is currently no universally accepted approach to identifying experts. As a result, this study will not rely solely on one method but rather use a combination of different approaches to assess whether someone is an expert. Prior to conducting expert interviews, the researcher has conducted brief initial calls with potential candidates. These calls assessed the candidate's experience, knowledge, and ability to meet key measures of discrimination and consistency as identified by Shanteau et al. (2002) These measures include the ability to perceive and act on subtle distinctions that others may miss, as well as the ability to consistently repeat judgments in similar situations. During this study, three potential participants initially perceived as experts were excluded. Through the course of the interviews, it surfaced that these individuals lacked the required expertise and experience. Consequently, their perspectives were not incorporated into the analysis.

Appendix C

This Appendix shows the interview protocol to understand the dimensions and factors of discrimination and the challenges of realizing a non-discriminatory financial lending system

Interview protocol

The interviews were conducted using a semi-structured interview protocol tailored to each interviewee, considering their expertise and background. This ensured that the questions asked were relevant and specific to each individual. It is important to note that each interview was unique and different from one another, as the interview protocol was tailored specifically to the interviewee. For instance, a data expert had different interview questions compared to a non-discrimination expert. However, the overarching objective of all the interviews remained consistent, which was to explore the dimensions, factors, and challenges associated discrimination.

Introduction [3 min]

- Welcome participant
- Provide an overview of the purpose of the interview
- Confirm consent and time expectation

Personal and professional background [5 min]

- Could you provide a brief overview of your professional background (e.g., years of experience, types of projects they worked on)

Dimensions of discrimination [10 min]

- How do you define discrimination?
- What is discrimination in your point of view?
- What is discrimination in financial lending context?

Factors of discrimination [30 min]

- What do you think is the main cause of discrimination in financial lending?
- What factors contributes to this discrimination?

Challenges [15 min]

- What are the challenges in realizing a non-discriminatory P2P lending/ financial lending system?
- Present challenges and inquire about their agreement when they are unable to recall additional ones.

Conclusion [3 min]

- Summarize the main conclusions of the interview
- Thank the interviewee

Discuss next step and the follow up evaluation

Appendix D

This Appendix summarizes the most important and interesting findings from the 13 interviews for sRQ 1 and sRQ 2.

A1:

A1 highlighted the challenge of addressing discrimination, as many of the cases are indirect and difficult to prove. They discussed that effectively combating discrimination requires legal measures. However, they also stress that legal measures alone are not enough. Addressing discrimination also includes addressing stereotypes by identifying them and showing positive aspects of disadvantaged groups, as well as addressing historical injustice and social inequality. Moreover, they recognized the significance of providing grants, quotas, and other support mechanisms to assist disadvantaged groups, instead of solely relying on individuals and corporations to solve the problem. They also highlighted the difficulty in identifying and proving indirect discrimination and the need for better legislation to tackle it, like GDPR. Moreover, discrimination is currently not given the same level of attention as other issues like privacy, due to a perception that it will cost money and not affect everyone in the same way. However, discrimination affects society, regardless of whether an individual belongs to the discriminated group or not, making it a matter of concern for everyone. To make combating discrimination a priority, it is essential to demonstrate its cost-effectiveness and raise awareness about its impact.

A2:

A2 is an experienced Dutch investor in P2P lending platforms, who has used several platforms for multiple years, but none of them were from the Netherlands. They expressed interest in using Dutch P2P platforms but noted that there are not many available and they do not know why. Their motivation for using P2P lending platforms is primarily to diversify his portfolio and achieve a good return on investments with reasonable risk, and they believe most lenders share this motivation. The potential returns on P2P loans are appealing, particularly when contrasted with the interest rates offered by traditional banks for savings accounts. They appreciate the flexibility and control that P2P platforms offer him, as they can choose which loans to invest in. At the same time, they acknowledged that they sometimes struggle to trust borrowers and will not provide loans to them in such cases. They have no conscious intention to discriminate when making these decisions, but rather relied on their own criteria to minimize risk. An interesting point they raised was that both the borrower and the P2P lending platform cannot check how they arrived at their lending decision. They stated a preference for lending to businesses over individuals because in their opinion there is less risk of default. They have been fortunate enough to only have experienced a defaulting loan once, but this did not discourage them from continuing to borrow through P2P platforms. They mentioned that the lender must make the judgment and therefore is responsible for the risk and consequences of decisions. They are a strong believer in P2P lending and its potential to disrupt traditional lending markets. They see it as a way to empower individuals, both from the borrower as well as lender side. It cuts out the trust required from a middleman like a financial institution and therefore reduces costs and overhead. They are optimistic about the future of P2P lending and believe it will be a growing market.

A3:

A3 emphasized the uniqueness of loans and the importance of financial inclusion for previously excluded populations. They noted that giving loans to those who cannot afford them can lead to greater financial losses, creating a complex balance between fairness, the customer well-being, and the best interests of the organization. According to A3, fairness is both context-specific and subjective. It should be considered alongside other competing priorities, such as customer well-being. The issue of algorithmic bias was discussed, with them explaining that even if certain features are removed, algorithms can still derive this information from other variables (proxies) due to societal inequalities. They argued that this is more of a problem with human cognitive biases, and that algorithms could be used to expose these biases. They pointed out that most scholarly research when it comes to biases focuses on the factors that influence credit decisions, such as borrower preferences with pictures on their profiles or connections to trusted networks. They also mentioned that research has shown that there is bias in the credit scores themselves, but that this is a separate issue related to credit risk bureaus rather than P2P lending. One of the biggest challenges in lending, especially P2P lending, is the problem of the "known unknowns", leading to the exclusion of marginalized populations. They also noted that a lack of accountability mechanisms or testing procedures could lead organizations to blame algorithms for biased results. They emphasized that pre- and post-processing should only be used in limited cases, as it can negatively impact a model's performance, explainability, and robustness. They argued that it is critical to address bias at its source, which may include training employees to reduce bias in data capture.

A4:
According to A4, institutional racism plays a role in the inconsistent distribution of money among Dutch entrepreneurs. This is partly due to personal biases in lending and a lack of role models for underrepresented groups. Underrepresented entrepreneurs also struggle to connect with the right investors and may encounter language barriers or biases during the application process, which can lead to feelings of discrimination and discouragement. Studies even found that female entrepreneurs are more successful than their male counterparts, contrary to stereotypes held in the financial sector. Despite claims of equal opportunities for all, investors are missing out on potential profits by excluding women from investment opportunities. This unconscious bias towards masculine qualities in successful entrepreneurs is deeply ingrained in the male-dominated financial sector. The financing industry in the Netherlands primarily supports two types of businesses: low-risk businesses with collateral and high-risk businesses that aim to quickly sell or merge quickly. Steadily growing family businesses, which generally have a more feminine approach to entrepreneurship, struggle to find funding. When designing a P2P lending platform in the Netherlands, it is important to consider the detailed European legislation on crowdfunding. Business loans may be a more relevant focus as consumer loans are almost non-existent in the Netherlands due to strict interest rate regulation making it unprofitable. Although crowdfunding involves consumer financing, the literature has yet to distinguish it from other forms of financing. Legal regulations make a clear distinction between non-accredited investors who borrow money as a private person under the European Crowdfunding Regulation if the transaction is aimed at small and medium-sized companies and less than a certain amount of money.

A5:
A5 provided valuable insights into discrimination law and highlighted the different types of discrimination, namely direct and indirect. Direct discrimination, in which banks or companies directly indicate that they will not offer a loan or job to a certain group because of the protected characteristics, is less common than indirect discrimination, which is often characterized by neutral criteria that lead to discrimination against certain groups. They also noted that discrimination law has become more complex, and the number of protected grounds is increasing. Each of these grounds has its own system and justifications, making it challenging for people to navigate the legal system. One of the main challenges in tackling discrimination is collecting sufficient evidence. The burden of proof is on the discriminated against, who must first identify the specific criteria used by the company. Gathering evidence requires contact with others in similar situations, which can be difficult. A preliminary level of evidence is needed before approaching the company for transparency. Companies are then obliged to provide clear information and avoid ambiguous language that could obscure the truth. A5 also stressed that discrimination is not only a legal problem, but a reflection of human behaviour and attitudes. As such, it cannot be fully regulated by law. In addition, the challenges in addressing discrimination often stem from emotional factors, as people are hesitant to complain about discrimination due to its sensitive nature.

A6:

A6 explained that in the Netherlands, concrete legislation for equal treatment is in place, based on international guidelines, which are implemented through Dutch legislation. The main law is the General Equal Treatment Act which prohibits making a distinction, rather than using the word discrimination, in the offering of goods and services such as loans. The law works with a fixed list of legal grounds for discrimination. Discrimination is defined as making a distinction based on irrelevant personal characteristics, but distinction on relevant characteristics such as a diploma is allowed. The law prohibits direct discrimination, while indirect discrimination is allowed with justified reason and is evaluated by the College for Human Rights. They mentioned as an example that exclusion based on skin colour, ethnicity, gender, or sexual orientation is unacceptable and unfair, but being denied a mortgage due to age is somewhat understandable as a borrower may not be able to pay it off before retiring. However, the decisions made by this College are not legally binding, unlike with privacy laws. The violation of privacy often intersects with discrimination, particularly when certain groups' privacy is violated. Furthermore, the legal framework for equal treatment is not comprehensive since topics like social class, which is a significant form of inequality, are not yet clearly protected by law. They emphasized the importance of being able to substantiate discrimination claims with international guidelines. Meanwhile, the word discrimination should not be used, as it can lead substantial resistance. They also highlighted that age discrimination is only prohibited in employment situations in the Netherlands, and that discrimination when offering products and services like banking and insurance based on age is not yet prohibited. They mentioned that the government is discriminating based on age when deciding on the minimal age to consume alcohol. Another interesting point they mentioned is that people have an unconscious bias towards people who resemble them and that this leads to unconscious discrimination. To address this, one might use more objective criteria and clear explanations to help filter this out.

A7:

A7 explained that it is essential to define goals and understand the user group and business objectives involved when addressing problems. Creating customer profiles and empathy maps helps determine user characteristics and needs. Factors like age, physical impairments, cultural backgrounds, and skin colour should be considered when segmenting user groups as they significantly impact user experiences. They emphasized that the typical persona created by commercial entities, however, always seem to focus on the people of whom they expect to make most profits. Governments invest in general accessibility, because it is their mandate, but commercial companies do not have this mandate. Pushing inclusivity issues onto the agenda can be challenging, especially if they impact timelines, budgets, or pose other obstacles. Companies tend to address such concerns if they do not significantly increase costs, complicate matters, or cause delays. Especially in the financial services sector, conservatism is prevalent, leading to a reactive rather than proactive approach. Innovation is typically adopted after competitors have proven its success. Regulatory requirements often drive companies to prioritize inclusivity due to the perceived downside risk of non-compliance and the desire to avoid financial losses or penalties. However, practical challenges and issues with prioritization can hinder the implementation of inclusive design, despite the existence of design systems that provide frameworks for addressing inclusivity. They mentioned that companies prioritize initiatives that maximize profits, which typically is at the expense of things that promote inclusion. Clear guidance on inclusivity is often lacking, but it is crucial to align inclusion with business goals from the start. By incorporating inclusivity considerations early on, segmenting user groups based on their unique experiences or goals, and embracing accessibility as an early adopter, companies can attract a significant user base. Despite the obstacles, inclusive design remains crucial for comprehensive and empathetic design practices.

A8:
A8 noted that when working with data sets, ensuring representativeness can be challenging. For instance, determining whether income data for specific consumer groups is representative requires careful consideration. The granularity level, category size, and geographical area play a role in verifying representativeness. Trusting externally bought data sets or internal statistical analysis may be necessary, but it is not always feasible to check representativeness at every step. Filtering out every potential risk of exclusion or discrimination is not possible, and it is crucial to be mindful of unintended biases in models and data used for analysis. They emphasized that understanding the financial industry comprises diverse individuals with different purposes and attention points is essential. However, accidents and unintentional exclusions can still occur due to the unknown nature of subgroups involved. Adjustments can be made if missing customer groups are identified, but without being informed of their absence, it becomes challenging to address the issue. They called these the "unknown unknowns", which are the biggest challenges. Known unknowns can and should be fixed. Regulation is playing a major role here. However, ethical considerations also come into play, as consciously making it easier or harder for specific groups reflects a moral decision during algorithm training or selection. While prioritizing certain customer groups may make sense from a profit perspective, it remains important to ensure that defined processes are followed consistently.

Architects play a vital role in establishing data pipelines and interfaces between applications. While their focus lies on implementing interfaces, the content flowing through them is often neglected. This approach poses risks as implementers or users determine the data that flows through the pipeline, potentially leading to improper implementation or excessive control that excludes certain information. It is crucial to incorporate feedback loops to identify potential discrimination and encourage inclusivity. Cultivating an inclusive culture during the reviewing process is essential, rather than removing information to mitigate discrimination risks. Transparent discussions should be encouraged to explain why certain cases receive more attention, and providing statistics on reviewed cases can enhance transparency, such as indicating the number of cases involving specific demographics. The pre-processing stage presents challenges as exclusions that occur during this phase are difficult to detect and rectify. They mentioned unknown unknowns again as the most significant challenge, where the existence of excluded groups is unknown. Furthermore, documentation should provide reasons for including or excluding specific data or conducting analyses, ensuring transparency in decision-making processes.

A9:
In the Netherlands, numerous strict rules and regulations are imposed by the AFM when it comes to providing loans to consumers. The intent behind these regulations is to protect financially vulnerable customers from taking on excessive debt or borrowing that they cannot afford. It is important to note that extending customer loans to customers with a higher default risk is usually highly profitable for banks because they factor that risk into the interest rates they charge. However, the AFM, as well as banks, strive to put consumer protection above profit maximization. Consequently, the unintended effect could be that some segments in the Netherlands have a challenge to access financing through banks. These segments may then revert to other informal or more obscure channels, which are less visible and potentially malicious, leading to even greater financial risks.

Interestingly, the Netherlands does not have multiple external credit bureaus, which is why the country primarily relies on one organization, the BKR, to calculate credit scores. As a result, people who are excluded from the credit score provided by this agency are also excluded from other organizations using the same score. Having multiple credit bureaus using different sources of credit information and methods to calculate credit scores could provide a more comprehensive evaluation of creditworthiness, benefiting those excluded from traditional assessments. Moreover, in the Netherlands, the focus is on not borrowing excessively and ensuring that loans are repaid, rather than improving the credit score as is the case in the United States. Nevertheless, individuals who do not fit into the Dutch loan system often experience difficulties in obtaining loans due to these regulations and credit rating practices.

A10:

A10 highlighted the importance of accessibility and crowdfunding as means of addressing financial needs. They discussed that regulators and banking supervisors now require lenders to justify loan approvals or denials based on objective criteria, rather than relying solely on subjective opinions. This shift towards objectivity can be observed in the corporate lending sector, where subjective models are being replaced with more objective ones. However, they noted that even objective models involve judgment calls and potential biases. For instance, certain individuals may be adversely affected by parameters in a mathematical model, leading to false negatives. One prominent indicator in loan applications is salary, as it demonstrates the borrower's ability to repay. Nevertheless, the persistent issue of gender pay gaps lies with employers, rather than the credit risk models themselves, and should be therefore tackled from there. Another is when proxies are used that can lead to discriminatory effects, such as migration background. While these proxies are based on objective data, they reflect discriminatory practices prevalent in the labour market. The legal context may not classify these practices as direct discrimination, but rather because of existing biases. Furthermore, in peer to peer the individuals decide themselves whether they want to lend out money, which makes it even harder to tackle. They also pointed to the limitations of anti-discrimination legislation and acknowledged the need for concrete measures beyond legal frameworks. In this context, the BKR was mentioned as playing a crucial role in maintaining individual credit records. However, it is important to note that the BKR's role primarily revolves around record-keeping rather than actively addressing discriminatory practices. While other entities could potentially provide similar services, the widespread usage and mandatory nature of the BKR make it challenging for competitors to enter the market. The BKR is in its current form provides a credit score like metric when financial organisations request a persons' data. They argued that this ideally should just be a list of open credits, which is more objective. They also acknowledged the possibility of subjectivity and data errors in the process. Requesting the removal of a registration requires a motivated reason, which may be influenced by subjective factors. Mistakes or unfavourable registrations can hinder individuals' access to funds, further underscoring the need for accurate and fair credit assessments.

A11:

A11 started by explaining the background check procedure which is often the initial step in the lending process. Creating an understanding of a financial situation involves various methods, such as analysing credit data, submitting payroll statements, and reviewing annual reports. When it comes to data privacy, financial organisations face the difficulty of assessing loan applicants' suitability and financial compatibility without requesting their entire life history. They mentioned that banks might use transaction data of customer, but many organizations will not have access to this data. Bias may potentially arise in how financial profiles are constructed, as a subset of information may be insufficient to fully represent an individual's circumstances. Biases in models can arise from two main factors. Firstly, if the training dataset does not offer a complete view of the entire population, biased outcomes can result. They gave an example of when a model is primarily trained on data from fast-food restaurants and then applied to a bakery, the bakery may receive an inaccurate score due to differing characteristics. Inevitably, certain data points will be missing regardless of the dataset size. Secondly, biases can emerge from the interpretation of observations. Cultural practices, such as sending salary contributions back to one's village, may be misconstrued as extravagant living, introducing bias into the evaluation process. Addressing these biases is challenging, as they often occur unconsciously and are rooted in an individual's perspective and interpretation. What may seem logical to one person may appear illogical to another, highlighting the subjective nature of biases.

They elaborated that ensuring equal treatment of similar cases poses another challenge due to time constraints and compliance-driven changes that financial organizations face. Priorities may shift, and certain plans may be continually postponed in favour of pressing compliance matters. Regulatory bodies also have limited capacity and focus on addressing issues that affect the majority or pose significant risks to the financial system. Negative consequences, such as fines or negative publicity, often drive prioritization efforts. They emphasized how addressing biases is complex since many biases are unconscious and deeply ingrained in individuals' behaviour. While biases can be conscious decisions for specific target groups, tackling unconscious biases affecting most of the population is more challenging. With regular observations and split-second decisions, it becomes difficult to trace back and reproduce the reasoning behind biases. However, adopting a factual approach can help mitigate biases and resolve related issues.

A12:
A12 discussed the importance of behavioural data in finance, but privacy restrictions and limited availability hinder research and applicability. They discussed the privacy fairness dilemma, where obtaining more personal information is necessary for measuring and improving fairness. Challenges arise in investigating correlations, such as between ethnicity and literacy levels, without direct access to relevant data. Unlike traditional credit rating agencies, P2P lending platforms often do not have the required data. Despite these challenges, they highlighted the importance of designing measurements to increase inclusivity and address difficulties in using risk models and algorithms. In cases where an algorithm cannot account for factors like literacy, human decision-makers can be employed to make fair decisions. Privacy laws and complexity further complicate the issue. While privacy laws emphasize the importance of accurate data, ensuring fairness may require invading someone's privacy. For example, the GDPR, specifically the Payment Services Directive 2 (PSD2) regulation, allows banks in the Netherlands and Europe to access behavioural data. However, privacy laws prohibit discrimination, making it challenging to assess whether biases exist within the data. Transparency, objectivity, and minimizing subjectivity in P2P lending platforms are essential for trust and mitigating bias. They emphasized that it is crucial for the lender to trust the P2P lending platform's proper determination of interest rates based on risk. That is why subjectivity should reside within the platform, as they can establish consistent standards for everyone. Subjectivity on the lender side, however, is beyond control. Therefore, implementing bias mitigating measures within the platform is the primary focus to minimize bias, as it is within their control. They further elaborated that the distinction between bias and discrimination is complex, and scientific research surpasses business contributions in addressing these concerns. Guidelines on bias are too vague and not practical since it is so dependent on the specific use case and context. While businesses advocate for fairness and equitable practices, implementing these principles in practice remains a complex task due to contextual and individual dependencies. Ultimately, trust, accurate risk assessment, and alternative solutions to limited data are crucial to solve.

Appendix E

This Appendix shows the variables lenders can discriminate on according to literature. It is important to note that not all variables could be found due to the time limit of this study. However, the articles selected were the most cited and are considered the most relevant. The dimension that contributes to discrimination is the provided hard and soft information on the platform for lenders to make their decision whether to provide loans to the borrower with certain characteristics. The overview is presented in Table E1.

Table E1: Overview of the variables and its impact lenders can discriminate on

Source	Objective	Type of information	Tested Variables	Impact
Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? (Iyer et al., 2012)	Evaluate whether lenders in P2P lending markets can use borrower information to infer creditworthiness and to examine the screening ability	Hard information	Loan amount Maximum Acceptable interest rate Loan period Income Credit information Home ownership Delinquency information	The higher the acceptable interest rate of borrower, the less credit worthy The more or higher the amount delinquent, the less credit worthy
		Soft information	Pictures Personal text descriptions (purpose of loan, goal, personal situation etc.)	The more loans requested previously, the more credit worthy
Effects of Borrower-Defined Conditions in the Online Peer-to-Peer Lending Market (Qiu et al., 2012)	Analyse the transaction data of PPDai, a P2P lending market provider in China, to determine how borrower-defined conditions affect successfulness of loan funding	Hard information	Loan amount Loan duration Repayment scheme Credit score Number of listings successful/unsuccessful Income Number of friends Amount bid by friends	The higher interest rate or smaller loan amount, the higher funding probability Personal information and social capital are the most important factors influencing funding outcome
		Soft information	Loan description Personal information	
Microfinance Decision Making: A Field Study of Prosocial Lending (Galak et al., 2011)	Investigate the characteristics of borrowers that encourage lending through Kiva, an online microfinance platform, and analyse the patterns in lender preferences	Hard information	Borrower group size Loan size Loan duration Time to get loan filled	Lenders prefer borrowers who share similarities in terms of gender, occupation, and first name initial
		Soft information	Gender Occupation First name initial	The larger the group of borrowers becomes, the less likely it is to receive loans
Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets (Ravina, 2019)	Study the impact of personal characteristics on loan decisions in an online lending market	Hard information	Loan amount Maximum interest rate Credit score Credit information Income range Employment length Delinquency information	Posting one or more pictures increases the chance of getting a loan A beauty of above average has a higher chance of getting a loan compared to neutral. For woman this chance increases more than for men
		Soft information	Employment Race Gender	

			Old/Young Overweight Religion Beauty Perceived trustworthiness Perceived Creditworthiness	Being overweight leads to an increased probability of getting a loan
What's in a Picture? Evidence of Discrimination from Prosper.com (Pope & Sydnor, 2011)	To examine racial disparities in P2P lending and determine if discrimination is taste-based or statistical	Hard information Soft information	Credit information Loan amount Maximum interest rate Debt to income ratio Picture based Adults, children, buildings, animals Gender Race Home ownership Purpose of loan Age group Happiness Weight Attractiveness	Reduced probability of loan success and higher interest rate required to pay against people without pictures and people with pictures of blacks, elderly, and unhappy people. On the other hand, there is positive discrimination of pictures of woman and with military involvement
Trust and Credit: The Role of Appearance in Peer-to-peer Lending (Duarte et al., 2012)	Examine the impact of appearance-based impressions on P2P lending and determine if trustworthiness affects the likelihood of loan funding and borrower default rates	Hard information Soft information	Credit information Delinquency information Maximum interest rate Loan amount Recovery in case of default Income level Fraction of complex and misspelled words Number of photographs Gender Couple Age categories Race Obesity House/Car/Business owner Perceived trustworthiness Attractiveness rating Perceived wealth Perceived willingness to pay Listing text indicator Listing start date Photograph indicator Endorsement indicator Friends' indicator Group member indicator Group leader indicator	Homeowners, college graduates and people perceived to be trustworthy have a higher probability of funding Perceived trustworthy people pay a lower interest rate than individuals that are perceived less trustworthy The expected return is higher for individuals perceived as trustworthy while it is lower for more attractive borrowers Borrowers who appear more trustworthy have indeed better credit scores and default rates than those who appear less trustworthy
Personal information in peer-to-peer loan applications: Is less more? (Prystav, 2016)	Examine the impact of information availability on investment behaviour in P2P lending platforms and determine how personal information, loan ratings, borrower liquidity, and loan purpose affect investment decisions.	Hard information Soft information	Loan amount Bank investment Loan duration Number of loans Age Loan purpose Gender Study semester Place of study Study program Experience with securities/ P2P lending	Non existential purpose loans, e.g., renovations, vacations etc., have a lower probability of successful funding When little information is available, herding behaviour became standard and projects with higher previous funding have a higher chance of getting funded

Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending (Emekter et al., 2015)	Analyse the characteristics and credit risk of loans in the online P2P lending market using data from the Lending Club and to determine the factors that affect loan performance	Hard information Soft information	Maximum interest rate Loan amount Credit information Income Total funded Home ownership Loan purpose	Lower credit grade leads to higher default risk thus higher interest rates requested by lenders
Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions (Herzenstein et al., 2011)	Investigate how identity claims constructed in narratives by borrowers' influence lender decisions and loan performance in the P2P lending market, and to determine the impact of the number and content of these claims on loan funding and repayment.	Hard information Soft information	Loan amount Percentage reduction of interest rate Loan performance Trustworthiness Success Hard working Economic hardship Moral Religious	Claiming more identities leads to increased likelihood of loan funding and lower interest rate Borrowers presenting themselves as moral or trustworthy have a higher chance of repayment because they strive to fulfil that image Borrowers using economic hardship identities often lack the ability or willingness to fulfil their loan payment
When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans (Gonzalez & Loureiro, 2014)	Study the effects of personal characteristics (perceived attractiveness, age and gender) of lenders and borrowers on online P2P lending decisions.	Soft information	Gender Attractiveness Capability Happiness Confidence Investment experience Image quality	Perceived age forms the basis of perceived experience and competence and leads to higher loan allocations when more mature and lower allocations when perceived young
Building Consumer-to-Consumer Trust in E-Finance Marketplaces: An Empirical Analysis (Greiner & Wang, 2010)	Investigate the trust-building mechanisms in P2P lending marketplaces using data from Prosper and determine the impact of various factors on trust behaviour	Hard information Soft information	Loan amount Maximum interest rate Loan duration Credit information Deb to income ratio Group rating Group leader reward rate Endorsements Description length Image availability	Borrowers with endorsements, a longer listing description and/or images have a higher percentage funded and can get more favourable loan terms
Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending (Ge et al., 2017)	Investigate the predictive power of self-disclosed social media information on borrowers' default probability in P2P lending and to identify social deterrence as a new mechanism that explains this predictive power	Hard information Soft information	Loan amount Interest rate Repayment terms Status of repayment Age Gender Education Marital status Verification method Social account disclosure	Borrowers who have disclosed their social media account information voluntarily are less likely to default on their loans Borrowers with a large social network or who posted lots of messages are less likely to default on their loans

Appendix F

This appendix shows all the challenges identified through literature and interviews. Table F1 presents all the 30 sources from which challenges have been identified. Table F2 presents all 72 identified challenges through interviews.

Table F1: All 30 sources from which challenges have been identified

SOURCE ID	AUTHOR	YEAR	TITLE	#CHALLENGES
P1	Bajracharya et al.	2023	Recent Advances in Algorithmic Biases and Fairness in Financial Services: A Survey	10
P2	Morse, A., Pence, K	2020	Technological Innovation and Discrimination in Household Finance	5
P3	Bertrand J., Weill L.	2021	Do algorithms discriminate against African Americans in lending?	4
P4	J Chen	2018	Fair lending needs explainable models for responsible recommendation	2
P5	Wu, B., Liu, Z., Gu, Q., Tsai, F.-S.	2023	Underdog mentality, identity discrimination and access to peer-to-peer lending market: Exploring effects of digital authentication	2
P6	Jiang J., Liu Y.-J., Lu R.	2020	Social heterogeneity and local bias in peer-to-peer lending – evidence from China	2
P7	Li Y., Ning Y., Liu R., Wu Y., Hui Wang W.	2020	Fairness of Classification Using Users? Social Relationships in Online Peer-To-Peer Lending	2
P8	Tian, G., Wu, W.	2023	Big data pricing in marketplace lending and price discrimination against repeat borrowers: Evidence from China	1
P9	El Annas, M., Benyacoub, B., Ouzineb, M.	2023	Semi-supervised adapted HMMs for P2P credit scoring systems with reject inference	1
P10	Sein, K.	2023	Crowdfunding credit services under the new proposal for a new Directive on Consumer Credits	1
P11	Caglayan, M., Talavera, O., Xiong, L.	2022	Female small business owners in China: Discouraged, not discriminated	1
P12	Emanuel-Correia, R., Duarte, F., Gama, A.P.M., Augusto, M.	2022	Does peer-to-peer crowdfunding boost refugee entrepreneurs?	1
P13	Wang T., Zhao S., Shen X.	2021	Why does regional information matter? evidence from peer-to-peer lending	1
P14	Chen X., Huang B., Ye D.	2020	Gender gap in peer-to-peer lending: Evidence from China	1
P15	Kim D.	2020	Sexism and ageism in a P2P lending market: Evidence from Korea	1
P16	Ayal S., Bar-Haim D., Ofir M.	2019	Behavioral biases in peer-to-peer (p2p) lending	1
P17	Fraser S.	2019	Entrepreneurial borrowing: Do entrepreneurs seek and receive enough credit?	1
P18	Chen D., Li X., Lai F.	2017	Gender discrimination in online peer-to-peer credit lending: evidence from a lending platform in China	1
P19	Kuwabara K., Thébaud S.	2017	When beauty doesn't pay: Gender and beauty biases in a peer-to-peer loan market	1
P20	Riggins F.J., Weber D.M.	2017	Information asymmetries and identification bias in P2P social microlending	1
P21	Lin M., Viswanathan S.	2016	Home bias in online investments: An empirical study of an online crowdfunding market	1
P22	Harkness S.K.	2016	Discrimination in Lending Markets: Status and the Intersections of Gender and Race	1
P23	Shi L., Zhang L.	2016	Regional discrimination in P2P lending of China	1
P24	Jenq C., Pan J., Theseira W.	2015	Beauty, weight, and skin color in charitable giving	1

P25	Loureiro Y.K., Gonzalez L.	2015	Competition against common sense: Insights on peer-to-peer lending as a tool to allay financial exclusion	1
P26	Barasinska N., Schäfer D.	2014	Is Crowdfunding Different? Evidence on the Relation between Gender and Funding Success from a German Peer-to-Peer Lending Platform	1
P27	Gonzalez L., Loureiro Y.K.	2014	When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans	1
P28	Chemin M., De Laat J.	2013	Can warm glow alleviate credit market failures? Evidence from online peer-to-peer lenders	1
P29	Dongyu C., Hao L., Xu H.	2013	Gender Discrimination towards Borrowers in Online P2P Lending	1
P30	Pope D.G., Sydnor J.R.	2011	What's in a picture?: Evidence of discrimination from Prosper.com	1

Table F2: All 72 identified challenges through interviews

CHALLENGE ID	CHALLENGE NAME	DESCRIPTION	INTERVIEWEE
C-IN1	Lender's subjective judgement	Subjective human judgments of the lenders can play a role in assessing business plan loans	A10
C-IN2	Balancing Competitive Advantage and Transparency	Financial organizations may have a competitive edge by using certain techniques over others, but they may not want to disclose what they are using publicly, leading to transparency issues in lending	A12
C-IN3	Anti-discrimination often not prioritized or postponed	Financial organizations are often constrained by time constraints and compliance-driven changes, which can lead to certain initiatives, such as anti-discrimination, not being prioritized or constantly postponed.	A11
C-IN4	Discrimination not receiving the same level of attention as other issues	Discrimination has not been given the same level of attention and prioritization as other important issues, such as privacy.	A1
C-IN5	Increasing importance placed on addressing bias	Increasing the level of importance placed by organizations on addressing biases in algorithms, data sets, and decision-making processes.	A3
C-IN6	Prioritization of organization targeting the most profitable group	Financial organizations prioritize targeting the most profitable group rather than ensuring inclusion of all groups	A7
C-IN7	Convincing organisation of benefits of accessibility	Convincing commercial companies that investing in accessibility from the beginning is not only a moral obligation, but also a sound business strategy.	A7
C-IN8	Convincing organisation of benefits of accessibility	Persuading organizations that proactively ensuring accessibility is more cost-effective than reacting to issues as they arise	A13
C-IN9	Motivation to reduce discrimination	Motivation to offer a form of modality that caters to the diverse needs and preferences of different groups of people.	A7
C-IN10	Prioritization of organisation based on ROI	Prioritization is often driven by the greatest return on investment, leaving certain groups behind	A7
C-IN11	Organisation rather reactive than proactive	Financial organizations are often reactive rather than proactive when it comes to tackling discrimination, only taking action if it becomes a legal obligation or if it hurts their bottom line	A7
C-IN12	Organisation rather reactive than proactive	Organizations are reactive rather than proactive when it comes to issues such as discrimination. This means that they only act if someone complains about it or hurt their reputation	A13
C-IN13	Organisation rather reactive than proactive	Organizations are reactive rather than proactive when it comes to issues such as discrimination. This means that they only act if someone complains about it or if it is legally required to do so.	A6
C-IN14	Limited resources to create inclusivity	Creating inclusive designs can be difficult, especially for smaller organizations with limited budgets.	A7
C-IN15	Balancing privacy and information disclosure	A lot of information is needed to properly assess someone's creditworthiness. However, many people are unwilling to disclose extensive personal information to a financial organization in order to obtain a loan.	A11

C-IN16	Privacy laws hinder discrimination detection	Privacy laws make it difficult to detect indirect discrimination because certain information cannot be collected or processed	A12
C-IN17	Balancing privacy and discrimination detection	Financial organizations may not have access to certain personal characteristics, such as ethnicity, due to privacy regulations. Therefore, it can be difficult for them to determine whether or not the loan results are related to these characteristics.	A12
C-IN18	Balancing privacy and fairness	Lending decisions involve a trade-off between privacy and fairness. Using personal data can make lending fair, but it can also invade individuals' privacy but not using these data can make it difficult to measure bias or fairness in the algorithm	A12
C-IN19	Missing ethnicity information to avoid discrimination	Using behavioral data in lending decisions can lead to biased results, as certain behaviors can correlate with a person's ethnicity, which is often removed from the data to avoid discrimination.	A12
C-IN20	Proxies can cause bias	The risk of bias does not lie in the obvious data points such as gender and nationality, which are typically filtered out, but rather in the use of proxies.	A10
C-IN21	Difficult to detect indirect discrimination	Indirect discrimination through proxies is difficult to detect and prevent and often results from discrimination earlier in the process, such as in the labor market.	A10
C-IN22	Proxies can lead to biased algorithms	Biased algorithmic decision-making can stem from the use of proxies	A12
C-IN23	Anonymization can hinder bias detection	Even if certain patterns are observed in a data set, identifying the underlying factors, such as gender or ethnicity, can be difficult due to anonymization practices. This information is often removed from the data before models are built, making it challenging to address the resulting biases in the model.	A11
C-IN24	Hard to track decisions leading to discrimination	When decisions are made by people without clear documentation, it can become difficult over time to explain the reasoning behind a decision that led to discrimination.	A11
C-IN25	Model Interpretation Can Introduce Bias	Interpretation of a model's output can introduce bias, even if it was not intentional. Interpreting results based on viewpoints can lead to subjective decisions and discrimination	A11
C-IN26	Subconscious Bias Can be Hard to Track	Biases are often subconscious or unconscious, making it difficult to discover where in the process that bias occurred.	A11
C-IN27	Biases Can Occur in Various Process Stages	Biases can creep in at various stages of the process, from data collection to algorithm design to human decision making, making them difficult to identify and address.	A12
C-IN28	Discrimination can occur in different stages	Discrimination can be introduced at any stages of the process making it difficult to identify the source	A8
C-IN29	Difficult to determine steps to discrimination	It can be difficult to determine the steps that led to discrimination, as they may not be transparent or documented	A8
C-IN30	Architects do not have overview of specific data flows	Architects often focus on designing the high-level architecture of the pipeline while leaving the specific data flows, which can introduce bias and discrimination, up to another team	A8
C-IN31	Preprocessing to Address Bias can Backfire	Preprocessing data to address biases can result in a dataset that is no longer representative of the population for which it is intended, leading to potential negative impacts on model performance and explainability.	A3
C-IN32	Preprocessing Can Lead to Invisible Discrimination	Preprocessing is often the most impactful and invisible form of discrimination. This is because if data is excluded or biased at this stage, it will never appear later in the process.	A8
C-IN33	Difficult for users to prove discrimination	Indirect discrimination is difficult to prove by users, which makes it harder to address.	A1
C-IN34	Difficult to prove discrimination due to lack of evidence	Discrimination is difficult to prove due to the lack of evidence	A3
C-IN35	Difficult to prove discrimination as organisation will deny	Discrimination is difficult to prove by users as organisations will deny	A13
C-IN36	Difficult to prove discrimination due to lack of evidence	Discrimination is difficult to prove due to the lack of evidence	A5

C-IN37	Difficult to prove algorithmic bias	When decisions are made by algorithms, it can be difficult for users to discover or prove whether they are discriminated by it	A6
C-IN38	Difficult to prove algorithmic bias	When decisions are made by algorithms, it can be difficult for users to discover or prove whether they are discriminated by it	A13
C-IN39	Difficult to speak up about discrimination	People do not like to complain or speak up about discrimination because it is a sensitive area	A5
C-IN40	No concrete guidelines on the prevention of discrimination	There are no strict guidelines for financial organizations to address the possibility of bias in the assessment of credit applications.	A11
C-IN41	Lack of guidance on preventing discrimination	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	A1
C-IN42	Lack of guidance on preventing discrimination	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	A7
C-IN43	Lack of guidance on preventing discrimination	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	A11
C-IN44	Lack of guidance on preventing discrimination	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	A6
C-IN45	Lack of guidance on preventing discrimination	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	A12
C-IN46	Balancing Fairness and Long-Term Welfare	Enforcing a fairness restriction on a lending platform could lead to greater inequalities between groups in the long run. Giving loans to people who can't afford them will result in defaults over time, lower credit scores, and greater differences in credit scores between the two groups.	A3
C-IN47	Relying on models to mitigate bias	Relying solely on a model to mitigate human biases can introduce new biases	A12
C-IN48	Human judgement leading to biases	Building a model involves subjective human judgments in determining which parameters to consider, potentially introducing biases into the model	A3
C-IN49	Human bias in data	Human biases can become embedded in data sets and affect the outcomes of algorithms	A3
C-IN50	Subjectivity exists when making decisions	When working with people there will be a subjective element involved in making decisions.	A12
C-IN51	Balance of subjective vs objective criteria	Balance the pros and cons of using subjective versus objective criteria in decision making. Objective criteria can provide more consistency, but can also lead to false negatives and potentially exclude qualified individuals. But subjective criteria can allow for more flexibility and account for individual circumstances, but they can also result in false negatives and false positives.	A10
C-IN52	Deeply rooted stereotypes	Gender bias can be deeply rooted in the financial industry, leading to missed opportunities and sustaining stereotypes that have been proven to be untrue.	A4
C-IN53	Bias is often unconscious	These biases is that they often happen unconsciously, based on that person own viewpoint and background	A11
C-IN54	Individuals are often unaware of their bias	It is difficult for individuals to be aware of their own biases, especially in a society where certain biases are normalized.	A12
C-IN55	Individuals are often unaware of their bias	People can find it difficult to be aware of their own biases, especially if they have not personally experienced discrimination	A2
C-IN56	Deeply rooted stereotypes	It is difficult to combat gender stereotypes in the financial sector, especially when it comes to investing in female entrepreneurs.	A4
C-IN57	Ensuring an inclusive mindset for decision makers	Ensuring that those responsible for making decisions and evaluating information have an inclusive mindset	A8
C-IN58	Presence of unknown unknowns	The presence of unknown unknowns, such as relying on algorithms based on past approvals, which can lead to lending only to similar people, further excluding populations that are already marginalized.	A3
C-IN59	Missing datapoints	Even with large datasets, there will be missing data points in models and analyses as you cannot collect everything	A11
C-IN60	Representativeness of dataset	The dataset used for model training may not provide a complete picture of the entire population, resulting in a biased dataset.	A11

C-IN61	Hard coded biases	Addressing unintended biases in algorithms that are hard-coded due to lack of diversity and representation in the development process.	A3
C-IN62	Assessing the Impact of Variables in Different Contexts	Understanding the impact (advantages and disadvantages) of including certain variables as input to algorithms in different context	A12
C-IN63	Assessing credit while considering context	Accurately assessing an individual's creditworthiness based solely on financial data is difficult without considering contextual factors	A11
C-IN64	Developing systems while considering cultural differences	Developing financial systems and models that take cultural differences into account, especially when it comes to transaction behaviour.	A11
C-IN65	Offering services while considering cultural requirements	Offering financial products that cater to specific cultural or religious requirements, such as the Islamic halal financing that prohibits the payment of interest.	A4
C-IN66	Predicting all possible effects in advance	It can be difficult to identify all subgroups that could be affected in advance, making it impossible to guarantee the prevention of all incidents.	A8
C-IN67	Preventing discrimination incidents	It is not possible to prevent all incidents, as it is often difficult to identify in advance all possible subgroups that could be affected.	A13
C-IN68	Detecting all sources of bias	Detecting and addressing all sources of bias in a model in advance, as there may be unknown underlying factors contributing to bias that are not known until implementation.	A11
C-IN69	Presence of unknown unknowns	There are unknown unknowns in the lending process, particularly when it comes to not knowing which groups could have been excluded because of unawareness of a groups existence	A8
C-IN70	Organisation lacking accountability mechanisms	Organizations lack accountability mechanisms for the algorithms they use, leading to the possibility of blaming the algorithm for unintended outcomes.	A3
C-IN71	Organisation lacking transparency and explainability	Organizations often lack transparency and explainability in their algorithms, leading to difficulty in understanding certain decisions and blaming the algorithm for negative outcomes.	A13
C-IN72	Organisation blaming the lender	On P2P lending platforms, the responsibility for final lending decisions lies with the individual lenders, potentially allowing the platform to avoid responsibility for any discriminatory outcomes.	A12

Appendix G

This Appendix shows the interview protocol to derive these principles and an evaluation protocol used to evaluate the formulation and content of each principle.

Interview protocol

The interviews were conducted using a semi-structured interview protocol, which included the following topics and example questions.

Introduction [3 min]

- Welcome participant
- Provide an overview of the purpose of the interview
- Confirm consent and time expectation

Personal and professional background [5 min]

- Could you provide a brief overview of your educational and professional background (e.g., years of experience, types of projects they worked on)

Definition and challenges of discrimination [15 min]

- Explain how discrimination is defined in this study
- Are you familiar with the discrimination challenges?
- Do you recognize some of these challenges <list of challenges>?

Principles and best practices [40 min]

- Is realizing non-discrimination one of your priority?
- How can we address these discrimination challenges <list of challenges>?
- Are there any existing principles or guidelines that can be used?
- Can you recommend any best practices?

Conclusion [3 min]

- Summarize the main conclusions of the interview
- Thank the interviewee
- Discuss next step and the follow up evaluation

Table G1 displays the table that is sent to interviewees through email prior to the interview. This table is intended to help them comprehend the challenges related to discrimination in advance of the interview.

Table G1: The challenges with a detailed description

ID	Challenge	Detailed description
1	Indirect discrimination through proxies is difficult to detect	Due to privacy regulations such as the GDPR, certain data points, such as someone's ethnicity, cannot be collected or processed. This means that when certain patterns are observed in a dataset, it is difficult to discover if it is related to ethnicity through proxies, such as zip code, as these obvious data points are filtered out.
2	Difficulty in discovering and tracing where discrimination or bias occurred throughout the process	The challenge is that when decisions are made without transparency or clear documentation, it can become difficult to explain the reasoning behind a decision that has led to a discriminatory outcome over time. This is because biases can creep into various process stages, from data collection to algorithm design to human decision making, making them difficult to identify and address. Preprocessing data can exclude or transform data in a way that it is not representative anymore. It also is often the most invisible form of discrimination. This is because if data is excluded or biased at this stage, it will never appear later in the process. Moreover, because bias is often unconscious, it is difficult to determine where in the process the bias originated.

3	Difficulty for borrowers to prove that there is discrimination	The challenge is that indirect discrimination is difficult to prove by borrowers, which makes it more difficult to address. This is because borrowers often do not complain or speak up about it, as it is a sensitive subject and there is often a lack of evidence. It is difficult to prove when an algorithm makes a decision that the algorithm has discriminated against that borrower due to the lack of transparency. In addition, companies will always deny that they have discriminated against that borrower. They often do not take the responsibility or blame it on the algorithm.
4	Organisations tend to react to incidents of discrimination rather than being proactive in preventing them	Discrimination does not receive the same attention as other issues such as privacy. This is because financial organizations are often affected by time constraints and compliance-driven changes, which can lead to certain initiatives not being prioritized or continually postponed. In addition, prioritization is often driven by the greatest return on investment, leaving certain groups behind. That is why organizations are reactive rather than proactive when it comes to issues such as discrimination. This means that they only act if it damages their reputation or business results, if someone complains about it, or if it is required by law to do so. However, it is important to note that organizations cannot react to incidents when incidents cannot be easily proven by borrowers.
5	Human involvement introduces biases throughout the process	The challenge is that when working with people there will always be a subjective element involved in making decisions. For example, building a model involves subjective human judgments in determining which parameters to consider, potentially introducing biases into the model. This can be influenced by the viewpoint, stereotypes, or culture of the employee involved.
6	Humans are not aware of their own biases	It is difficult for individuals to be aware of their own prejudices, especially in a society where certain biases are normalized and not everyone is equally affected by them. Biases often occur unconsciously and are shaped by an individual's point of view and background. Gender stereotypes are also rooted in the financial sector, making this issue even more challenging
7	There is a lack of concrete guidance on how to ensure that a system or service is not discriminatory.	Currently, there is a lack of generally accepted and specific guidelines or standards for implementing anti-discrimination laws in practice. Similarly, financial organizations do not have concrete guidelines for designing systems that are free from discrimination from the beginning. In contrast, principles around the topic of privacy such as PbD and the GDPR are well established and provide concrete guidelines for organizations to follow.
8	Contextual factors can be overlooked by systems	Designing financial systems that take borrower's contexts into account can be difficult when cultural practices are for example not well understood. Models can label borrowers as high-risk based solely on their spending habits, regardless of cultural context. Some cultures dictate a monthly income to their family or village, leading to above-average spending patterns that can be mistakenly interpreted as risky.
9	Difficult to prevent all incidents due to unknown factors	Detecting and addressing all sources of bias in a model in advance is difficult, as there may be unknown underlying factors contributing to bias that are not known until implementation. This also means that it can be difficult to identify all subgroups that could be affected in advance, making it impossible to guarantee the prevention of all incidents.
10	Financial organisations tend to avoid responsibility and accountability	Organizations currently lack accountability mechanisms to address discriminatory outcomes resulting from algorithms. This often leads to the possibility of blaming the algorithm for unintended outcomes. Moreover, they also often lack transparency and explainability in their algorithms. On P2P lending platforms, the responsibility for final lending decisions lies with the individual lenders, potentially allowing the platform to avoid responsibility for any discriminatory outcomes.
11	Solutions to reduce bias and discrimination can have unintended negative effects	Solutions to mitigate discrimination might have unintended negative consequences. These may not become apparent until after they have been implemented. Even models that are designed to reduce bias can sometimes introduce new forms of bias. For instance, providing lenders with more information to help them reduce bias might end up validating the biases they already have. Additionally, providing loans to individuals who cannot afford them may seem like a solution to discrimination, but it can ultimately harm these individuals in the long run.
12	Bias can be introduced through data in many ways	Bias in data can arise from missing data points, incomplete data, reliance on data from specific groups, faulty data, skewed data, or by using selective variables. These might not be a good representation and therefore introduce bias. One of the causes is the unknown unknowns, such as relying on past approvals data leading to lending only to similar people. Furthermore, bias might be hard-coded in data due to lack of diversity in the development process.
13	Lenders can discriminate based on the provided information by borrowers and platform	Lenders can access various information about the borrower, including age, ethnicity, geographic location, and appearance. While this information is helpful in making decisions by increasing trust, it also presents a potential for discrimination. Lenders relying on such information may (un)intentionally make biased decisions, leading to unequal treatment of borrowers from certain groups.

Evaluation form

Introduction

Thank you for taking the time to evaluate the developed principles. Combining your insights and those of other architects, a set of 13 principles has been developed. These principles have been intentionally developed to provide high-level guidance that can be applicable to all financial organizations.

The survey is divided into two parts:

- The first part involves evaluating the formulation and content of each principle and providing feedback on aspects such as the name, statement, rationale, and implications.
- The second part involves ranking the principles based on their perceived impact and usefulness, from most impactful to least impactful.

Thank you again for your time and contribution.

General information

What is your name

Please note that I will not use your name in my research; this is solely to understand who I should approach in case I need further clarification on an answer.

[ ]

Principles Evaluation (only one principle is showed as an example)

Name	Dedicated Ethics Team
Statement	Establish an ethics team within the organization to proactively address discriminatory topics and uphold non-discrimination principles throughout the lending process
Rationale	A dedicated team helps to ensure that discrimination topics are not deprioritized or overlooked.
Implications	• Allocate resources for the establishment and operation of the ethics teams/ officer • Establish collaboration and clear communication channels between the ethics team/ officer and other teams • Provide the team with authority and resources to effectively address potential discriminatory issues • Foster a culture of openness and continuous learning

Are there any comments or suggestions regarding the formulation and content of this principle?

Name, statement, rationale, and implications

[ ]

Principles Ranking

Rank the principles based on their usefulness to realize non-discriminatory financial systems. From most useful (rank 1) to least useful (rank 13)

[Row with principles
Column with ranking numbers]

Appendix H

This Appendix summarizes the most important and interesting findings provided by the interviewees to address the challenges of discrimination within financial lending systems for sRQ 3.

B1

- The presence of information can enable the prevention of discrimination while its absence can also serve as a safeguard against discriminatory treatment by individuals or algorithms.
- Analyse all business rules, considering different categories like gender and ethnicity discrimination. To illustrate, evaluate each rule to understand its impact on individuals from diverse backgrounds.
- The principle of the "four eyes" can be employed to ensure that a process or decision is reviewed by at least two individuals. However, this approach can sometimes introduce additional biases. To address this, an alternative solution is to implement a "Round Robin" system or involve two randomly selected individuals who are unaware of previous checks conducted.
- Train an algorithm to identify uncertain situations related to discriminatory complaints or problematic data. By understanding the combination of business rules that contribute to such situations, the algorithm can be programmed to raise alerts instead of making definitive decisions. The aim is to prompt human intervention and evaluation in these uncertain cases.
- Have a neutral third party or licensed algorithm that does have all the privacy-sensitive data to conduct regular spot checks on several requests. During these spot checks, the algorithm's decisions or outcomes would be compared with a neutral third party or licensed algorithm to determine if they align.
- AI models can be used to oversee and monitor the performance of other models, providing an additional layer of control.
- Establishing a dedicated department or team, like a risk department, focused on preventing discrimination can be beneficial. This specialized unit would proactively address potential discrimination issues rather than reacting to them after they occur.
- Organisations can undergo a specific testing procedure to validate their business rules. This process involves testing that the collected data and business rules meet the standards of non-discrimination. Organizations can provide then provide proof of their commitment to non-discrimination.

B2

- When working with a system that incorporates multiple business rules, it is crucial to isolate changes by introducing or modifying one business rule at a time. This approach allows for a clear understanding of the specific implications and consequences associated with each individual change.
- Foster close collaboration between the team responsible for building the system and the business stakeholders. While the technical team plays a significant role in constructing the system and implementing business rules, it is essential to incorporate the business perspective and align closely between business and IT specialists.
- Encourage a collaborative decision-making process. Establish a strong partnership between business and IT, ensuring that the technical implementation aligns with the intended business requirements and design.
- Consider implementing a decision-making triangle, a collaborative framework that involves IT delivery, business stakeholders, and architectural teams. This framework facilitates effective communication and coordination among all parties involved in the development and implementation process.

- Emphasize the importance of the top layer of the organization making discrimination a priority or strategic objective. By setting a clear mandate and commitment to combating discrimination, it motivates all layers of the organization to take is seriously into account and do something about it.
- Financial organizations lacking historical data for certain groups can use a predictive credit risk scoring algorithm based on business rules. They analyze successful lenders from a limited dataset and build a prognosis model. Despite being initially deemed high-risk, these individuals later become successful lenders. The system then incorporates new data for accurate credit score estimation for similar groups.

B3

- Foster diversity in lending teams by incorporating members from various backgrounds, including gender, race, economic background, and relevant professions such as ethics specialists or sociologists. This diversity can bring different perspectives and insights to mitigate biases in decision-making processes.
- Provide ongoing education and training programs for employees based on research, factual data, and real-life cases to increase awareness of potential biases and empower them with strategies to address and mitigate biases effectively.
- Continuously assess and improve the balance between human judgment and the use of AI in decision-making processes. Identify situations where AI can help prevent biases and cases where AI itself may introduce bias, ensuring appropriate utilization of AI technologies.
- Implement "Embedded Ethics" by incorporating discrimination controls and testing them at all stages of the lending process, including advertising, pre-application, loan approval, and loan management. These controls should encompass both automated mechanisms and procedural guidelines to proactively address and prevent discrimination.
- Provide specialized training and education to developers and system designers to raise awareness of the potential for bias and discrimination in their systems. Equip them with the necessary tools and techniques to mitigate these risks during the system's development and design stages.
- Implement robust data governance processes to ensure the accuracy, transparency, and unbiased nature of the data used to train the lending system or service. Regularly review and update data sets to maintain diversity and representation, ensuring they remain accurate and up to date.
- Utilize "Explainable AI" techniques to enhance transparency and understandability of AI models. This enables stakeholders to comprehend how decisions are made and identify potential biases.
- Employ Bias Analysis and reveal tools, when available, to systematically evaluate and identify biases in the decision-making process. These tools can provide insights and recommendations for addressing and mitigating biases effectively.
- Leverage IT tools and technologies to ensure the accuracy, transparency, and representativeness of the data used in the system. Implement mechanisms for regular review, validation, and updates of data sets to maintain their quality and mitigate potential biases.

B4

- Establish guidelines and principles in advance for creating a non-discriminatory system. Test and validate these principles to ensure they are practical and effective in achieving non-discriminatory results.
- Include non-discriminatory considerations from the beginning of the development process, particularly during problem definition and goal setting. Considering non-discrimination as important makes it an integral part of the design and implementation of the system.

- Develop non-discriminatory acceptance criteria that go beyond mere documentation. These criteria must be well defined and verifiable, so that it can be objectively assessed whether the system complies with the non-discrimination criteria.
- Take immediate and direct action when discrimination is discovered, rather than allowing it to linger in the development backlog. Prioritize addressing discriminatory issues and document the decision-making process surrounding them. If a conscious decision is made not to address discrimination, it must be clearly documented.
- Recognize that transparency alone is not enough. It is essential that employees understand and can explain how the model works. Moreover, transparent models and acknowledging the presence of discrimination is pointless if no action is taken to deal with it.
- Establishing effective governance is essential to address discriminatory issues within an organization. Frontline employees should flag problems, which are then escalated to management and executives if necessary. The first line of defense resolves issues, but an escalation path ensures problems are addressed at higher levels if needed. Each layer of management is accountable for addressing discrimination.

B5

- Evaluate business rules and parameters to assess their potential for leading to discriminatory outcomes. It is crucial to understand the impact of these rules before incorporating them into decision-making processes.
- Obtain explicit permission from customers if access to sensitive data is necessary. It is important to ensure that data usage aligns with legal and ethical standards.
- Generate and test different hypotheses using available data to examine potential discriminatory effects. This approach helps uncover any biases or unfair outcomes associated with the algorithms or models.
- Prioritize the use of additional data sets beyond the primary ones to evaluate the performance of the algorithms in terms of detecting discrimination.
- Employ models to support decision-making, as they can be more objective and less prone to biases compared to humans. However, supplementing them with clear, objective guidelines for human decision-makers can help mitigate biases
- Establish an ethics department or team responsible for monitoring and addressing potential discrimination issues. They can then conduct regular tests on data and grant the department the authority to suggest and, if possible, implement procedural changes to mitigate discrimination.
- Make non-discriminatory performance a goal for every layer of the organization to ensure accountability

B6

- Enhance transparency by providing detailed explanations of the underlying processes and criteria used for loan approval, both internally and to customers. This will help individuals understand how decisions are made and identify any potential discrimination.
- Acknowledge that companies are profit-driven and highlight the need to align profitability with fairness. Emphasize the importance of regularly evaluating and adjusting the model to ensure it does not discriminate, rather than relying solely on profit optimization.
- Advocate for regulatory mandates that require companies to address discrimination in their models. Making it a legal requirement will increase motivation and accountability for combating discrimination.
- Consider involving a neutral third party to oversee and audit the process to ensure fairness. This independent entity can provide unbiased assessments of the model's behavior and help maintain transparency.

- Implement data lineage practices to track and document the flow of data throughout the process. This will facilitate tracing back decisions, ensuring accountability, and identifying any potential biases introduced at various stages.
- Automate processes as much as possible to enhance traceability and transparency. This will minimize manual intervention and reduce the chances of human bias influencing decisions.
- Foster a culture of openness and encourage reporting of issues. Promote an environment where individuals feel safe to raise concerns about discrimination and biases in the model. This will help organizations address problems effectively.
- Establish a framework that outlines specific checks and consequences for evaluating business rules and their impact. This framework should include mechanisms to assess and mitigate potential discriminatory effects, ensuring a fair decision-making process.

B7

- Ensure that the model's rules and calculations are traceable and explainable, even if certain parameters can change over time. The ability to understand how a certain result was reached is crucial for transparency and accountability.
- Define the level of explainability required based on the intended audience. Technical understanding is necessary for those working directly with the models, while those applying the models may not require understanding all the deep technical workings.
- Consider that justified discrimination, based on relevant factors in risk assessment, may be accepted by individuals. For example, if a risk model accounts for differences in life expectancy between genders and adjusts premiums accordingly, it may be considered fair treatment.
- Create awareness and understanding of discrimination at all levels of the organization. Establish a clear business goal that prioritizes addressing discrimination and translate it into actionable solutions.
- Ensure that people within the organisation also have the capabilities, tools and processes in place to resolve these discrimination issues.
- Assign responsibility to a dedicated team or integrate discrimination-related responsibilities into an existing team such as the privacy team. This team will be responsible for addressing and resolving issues related to discrimination.

Appendix I

This appendix shows the outcome of the initial evaluation of the developed principles by the architects who developed them. This includes what has been changed about the original principle leading to the final principles that will be re-evaluated by another group of architects on their usefulness.

Thus the principles has been changed based on the feedback from the evaluation form and the additional conversation with them for clarification and evaluation after they completed that form. GREEN is what added and adjusted; and RED is what is removed.

P1	Name	Dedicated Ethics Team or Officer
	Statement	Establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process.
	Rationale	A dedicated team or officer helps to ensure that discrimination topics are not deprioritized or overlooked.
	Implications	• Allocate resources for the establishment and operation of the ethics teams/ officer • Establish collaboration and clear communication channels between the ethics team/ officer and other teams • Provide the team/ officer with authority and resources to effectively address potential discriminatory issues • Removed: foster a culture of openness and continuous learning

P2	Name	Involve Diverse Perspectives in Collective Decision-Making
	Statement	Involve diverse perspectives in a collective decision-making process to ensure that no decision is made solely by a single individual or viewpoint.
	Rationale	The organization reduces the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying on a single individual and viewpoint.
	Implications	• Allocate resources and time for additional individuals' involvement • Clearly assign responsibility and establish accountability for decision-making outcomes • Anticipate potential delays due to the additional review step • Establish an escalation process for disagreements or conflicting viewpoints • Encourage active participation from individuals with different perspectives • Removed: Maintain proper documentation of the decision process • Ensure that decisions are backed by data when a second decision-maker is not available • Manage communication and coordination among team members

P3	Name	Evaluate the Models and their Documentation explicitly for Discriminatory Aspects and Document Business Rules Explicitly for Discrimination
	Statement	All elements of the model (e.g., data, algorithm, parameters) used business rules applied in the lending processes should be regularly evaluated and documented for potential discriminatory effects.
	Rationale	Regular evaluation of models business rules enables organizations to identify and resolve potential issues and gain timely insight into the impact of each rule. Documentation facilitates easy tracing and justification of the model these rules when required in the future.
	Implications	• Establish a guideline to assess the implications and alignment of business rules models with non-discrimination objectives • Keep models rules up to date with newest insights on discrimination • Maintain transparency and traceability in evaluation processes • Provide training to employees involved in testing and evaluation

P4	Name	Independent Oversight
	Statement	Have an independent entity that is responsible for overseeing and ensuring compliance to non-discrimination guidelines.
	Rationale	The introduction of independent oversight adds an unbiased layer of evaluation and control. It enforces accountability and resolves potential conflict of interests.
	Implications	• Establish relationships with independent entities to oversee and ensure compliance • Improve and create processes to support the work of independent entities • Be open-minded and prepared to act on their findings and recommendations, which may require changes to existing processes or practices • Regular reviews may incur additional costs

P5	Name	Create Awareness and Educate the workforce on bias
	Statement	Continuously train and educate all employees to recognize and address biases.
	Rationale	Educating employees to raise awareness about their biases is a valuable first step, as many are unaware of their biases and lack the knowledge to address them effectively.
	Implications	• Mandate awareness training for all employees, regardless of their role or level • Evaluate and update training effectiveness and relevance regularly • Allocate resources for providing training sessions • Establish the required organizational culture and mindset • Increased time and effort required for employees to participate in mandatory trainings

P6	Name	Have an Escalation Path and a Non-Retaliation Policy
	Statement	Establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy
	Rationale	Discrimination concerns and issues are not always properly raised, investigated or resolved. A clearly defined escalation path creates a safe environment for employees to report and address discrimination issues while demonstrating the organization's commitment to non-discrimination
	Implications	• Establish a structured process for raising concerns about discrimination at the appropriate level for timely resolution • Empower employees to report discriminatory issues without fear of retribution • Ensure that discriminatory issues are addressed and not ignored • Promote a culture of continuous improvement by capturing and applying lessons learned from resolving discriminatory issues

P7	Name	Culture of Openness and Learning
	Statement	Fostering an open culture where individuals are encouraged and incentivized to report discriminatory issues and learn from mistakes
	Rationale	For an effective escalation path, Individuals need to feel safe and encouraged to actively participate in identifying, reporting, and addressing discrimination. This promotes continuous improvement and strengthens the collective commitment to a non-discriminatory environment and process.
	Implications	• Have processes in place that encourages Foster a safe and open environment for individuals to speak up • Establish accessible communication channels and feedback mechanisms for reporting discrimination • Create incentives to encourage reporting and active participation in reducing discrimination • Clearly assign responsibility and establish accountability for the culture creation • Commit management and leadership to support foster this culture and lead by example • Update policies, processes and practices based on lessons learned • Have an escalation path in place that creates a safe environment for employees to report and address discrimination issues

It is important to note that Document for Traceability and Explainability in Lending process are merged into Traceable and Explainable Lending Process

P8	Name	Traceable and Explainable Lending Process Document for Traceability / Explainability in Lending Process
	Statement	All decisions of the lending process should be traceable and both internally understandable by employees and externally by customers and other stakeholders. Ensure traceability by documenting decisions at each decision-making stage.
	Rationale	Ensuring traceability, achieved through documentation and the ability to explain the tools, models, and processes utilized, is important as it enables transparency, Clear documentation enables traceability, understanding of decision-making processes, facilitates independent evaluations, and fosters accountability.
	Implications	• Choose and design tools, models and systems to be understandable by employees • Provide training for employees especially developers to understand and explain models to non-technical users • Design systems to trace decisions back to criteria and data

		• Ensure an accessible and secure system for storing and managing documentation • Implement procedures on what exactly and how it should be documented • Dedicate time and effort from employees as they need to consistently document decisions • Apply regular checks to ensure correct documentation

P9	Name	Ensure Representative and Relevant use of Data Data is representative and relevant
	Statement	Utilize data that is representative of the population and relevant to the lending processes
	Rationale	Organizations needs representative and relevant data to make well-informed and fair lending decisions.
	Implications	• Avoid relying solely on historical data • Consider collecting additional data to increase representativeness • Regularly review and validate data sources and collection methods

P10	Name	Work towards Undisputed Decisions Certainty in Decision-Making
	Statement	Strive for undisputed decisions and actively seek for additional input when faced with disputes. Decisions should be made with a high level of certainty, and additional input should be sought when faced with uncertainty
	Rationale	This helps to prevent potentially biased or discriminatory outcomes that could have been avoided
	Implications	• Delay disputed uncertain decisions until there is more consensus certainty regarding the decision at hand • Define clear process in case a decision is disputed uncertain • Include and use indicators that measure the accuracy certainty of a system or model's output • Train algorithms and employees to identify uncertainty

P11	Name	Monitor Continuously
	Statement	Discrimination, both direct and indirect, should be monitored checked at every stage of the process.
	Rationale	Continuous monitoring for discrimination helps to identify and mitigate biases or discriminatory practices that may emerge during the process.
	Implications	• Establish a comprehensive continuous monitoring process • Ensure monitoring mechanisms detect both direct and indirect discriminatory factors • Provide additional resources for monitoring • Report, document and communicate monitoring findings • Update policies, processes and practices based on findings

P12	Name	Non-Discrimination as an Objective
	Statement	Set clear non-discrimination objectives from the beginning and establish predefined acceptance criteria that align with these objectives.
	Rationale	By establishing acceptance criteria from the outset, non-discrimination can be recognized as a goal rather than merely a financial burden.
	Implications	• Define clear objectives and acceptance criteria in advance, including thresholds • Utilize metrics and KPIs to track progress • Assess performance against objectives regularly • Equip decision-makers with the capabilities to effectively achieve these objectives • Continuously review and update criteria to ensure relevance

Appendix J

This Appendix shows the workshop protocol to evaluate the usability, impact, formulation and content of each principle.

Workshop protocol
The workshop was an in interactive session where 12 principles were presented and evaluated, which included the following topics and example questions.

Introduction [3 min]

- Welcome participant
- Provide an overview of the purpose of the interview
- Confirm consent and time expectation

Personal and professional background [5 min]

- Could you provide a brief overview of your educational and professional background (e.g., years of experience, types of projects they worked on)

Definition and challenges of discrimination [15 min]

- Explain how discrimination is defined in this study
- Are you familiar with the discrimination challenges?
- Did you recognize some of these challenges and principles shared via email?

Principle evaluation [80 min]
Present each principle on the screen and start the evaluation discussion

- What do you think of the usability of principle [] for the improvement of non-discrimination in financial lending?
- What do you think of principle [] in terms of importance and why?
- Is principle [] easy to implement in practice and why?

Additional remarks [10 min]

- Are there any additional feedback and suggestions to improve these principles?

Conclusion [3 min]

- Summarize the main conclusions of the workshop
- Thank the interviewee
- Discuss next step

Appendix K

This appendix shows what has changed based on the evaluation workshop. GREEN is what added and adjusted; and RED is what is removed.

Name	Dedicated Ethics Team or Officer
Statement	Establish an ethics team or officer within the organization who is responsible for proactively addressing discriminatory issues and enforcing non-discrimination principles throughout the lending process.
Rationale	A dedicated team or officer helps to ensure that discrimination topics are not deprioritized or overlooked.
Implications	• Allocate resources for the establishment and operation of the ethics teams/ officer • Establish collaboration and clear communication channels between the ethics team/ officer and other teams • Provide the team/ officer with authority and resources to effectively address potential discriminatory issues • Ensure that the existing ethics team and officer expand their scope to address discrimination in the lending process, in addition to their current responsibilities

Name	Involve Diverse Perspectives in Collective Decision-Making
Statement	Involve diverse perspectives in a collective decision-making for important and impactful decisions to ensure that no decision is made solely by a single individual or viewpoint.
Rationale	The organization reduces the risk of bias, errors, discriminatory practices, and conflicts of interest that can arise from relying on a single individual and viewpoint.
Implications	• Allocate resources and time for additional individuals' involvement • Clearly assign responsibility and establish accountability for decision-making outcomes • Establish an escalation process for disagreements or conflicting viewpoints • Encourage active participation from individuals with different perspectives • Ensure that decisions are backed by data when a second decision-maker is not available • Manage communication and coordination among team members • Ensure that there are diverse perspectives within the organization • Introduce an additional review step specifically for areas that are highly vulnerable to discrimination

Name	Evaluate the Models and their Documentation explicitly for Discriminatory Aspects
Statement	All elements of the model (e.g., data, algorithm, parameters) used in the lending processes should be regularly evaluated and documented for potential discriminatory effects.
Rationale	Regular evaluation of models enables organizations to identify and resolve potential issues and gain timely insight into the impact. Documentation facilitates easy tracing and justification of the model when required in the future.
Implications	• Establish a guideline to assess the implications and alignment of models with non-discrimination objectives • Keep models up to date with newest insights on discrimination • Maintain transparency and traceability in model and evaluation processes • Provide training to employees involved in testing and evaluation • Change or remove parameters from the model that are discriminatory

Name	Independent Oversight
Statement	Have an external or internal independent entity that is responsible for overseeing and ensuring compliance to non-discrimination guidelines.
Rationale	The introduction of independent oversight adds an unbiased layer of evaluation and control. It enforces accountability and resolves potential conflict of interests.
Implications	• Establish relationships with external or internal independent entity to oversee and ensure compliance • Improve and create processes to support the work of independent entities • Be open-minded and prepared to act on their findings and recommendations, which may require changes to existing processes or practices • Regular reviews may incur additional costs

	• Develop an assessment framework to assess against

Name	Create Awareness and Educate the Workforce
Statement	Continuously train and educate all employees to recognize and address biases.
Rationale	Educating employees to raise awareness about their biases is a valuable first step, as many are unaware of their biases and lack the knowledge to address them effectively.
Implications	• Mandate awareness training for all employees, regardless of their role or level • Tailor the trainings to the specific roles and responsibilities of employees • Evaluate and update training effectiveness and relevance regularly • Allocate resources for providing training sessions • Establish the required organizational culture and mindset • Increased time and effort required for employees to participate in mandatory trainings

Name	Have an Escalation Path for Employees and Customers and a Non-retaliation Policy
Statement	Establish a clearly defined and documented escalation path for employees and customers to easily find and follow to resolve and address discrimination issues, coupled with a non-retaliation policy
Rationale	Discrimination concerns and issues are not always properly raised, investigated or resolved. A clearly defined escalation path creates a safe environment for employees to report and address discrimination issues while demonstrating the organization's commitment to non-discrimination
Implications	• Implement a non-retaliation which can also be achieved by leveraging existing non-retaliation policies if they are already in place. • Establish a structured process for raising concerns about discrimination at the appropriate level for timely resolution • Empower employees to report discriminatory issues without fear of retribution • Ensure that discriminatory issues are addressed and not ignored • Promote a culture of continuous improvement by capturing and applying lessons learned from resolving discriminatory issues.

Name	Culture of openness and learning
Statement	Fostering an open culture where individuals are encouraged and incentivized to discuss discriminatory issues and learn from mistakes
Rationale	Individuals need to feel safe and encouraged to actively participate in identifying, discussing, and addressing discrimination. This promotes continuous improvement and strengthens the collective commitment to a non-discriminatory environment and process.
Implications	• Have processes in place that encourages a safe and open environment for individuals to speak up • Establish accessible communication channels and feedback mechanisms for discussing discrimination • Create incentives to encourage discussions and active participation in reducing discrimination • Clearly assign responsibility and establish accountability for the culture creation • Commit management and leadership to support this culture and lead by example • Update policies, processes and practices based on lessons learned • Have an escalation path in place that creates a safe environment for employees to discuss and address discrimination issues

Name	Traceable and Explainable Lending Process
Statement	All decisions of the lending process should be traceable and both internally understandable by employees and externally by customers and other stakeholders.
Rationale	Ensuring traceability, achieved through documentation and the ability to explain the tools, models, and processes utilized, is important as it enables transparency, facilitates independent evaluations, and fosters accountability.
Implications	• Choose and design tools, models and systems to be understandable by employees • Provide training for employees especially developers to understand and explain models to non-technical users • Design systems to trace decisions back to criteria and data • Ensure an accessible and secure system for storing and managing documentation • Implement procedures on what exactly and how it should be documented

	• Apply regular checks to ensure correct documentation	

Name	Ensure Representative and Relevant use of Data
Statement	Utilize data that is representative of the population and relevant to the lending processes
Rationale	Organizations needs representative and relevant data to make well-informed and fair lending decisions.
Implications	• Avoid relying solely on historical data • Consider collecting additional data to increase representativeness • Regularly review and validate data sources and collection methods

Removed: Name	Work towards Undisputed Decisions
Statement	Strive for undisputed decisions and actively seek for additional input when faced with disputes.
Rationale	This helps to prevent potentially biased or discriminatory outcomes that could have been avoided
Implications	• Delay disputed decisions until there is more consensus regarding the decision at hand • Define clear process in case a decision is disputed • Include and use indicators that measure the accuracy of a system or model's output • Train algorithms and employees to identify uncertainty

Name	Monitor Continuously
Statement	Discrimination, both direct and indirect, should be monitored at every stage of the process.
Rationale	Continuous monitoring for discrimination helps to identify and mitigate biases or discriminatory practices that may emerge during the process.
Implications	• Establish a comprehensive continuous monitoring process • Ensure monitoring mechanisms detect both direct and indirect discriminatory factors • Provide additional resources for monitoring • Report, document and communicate monitoring findings • Update policies, processes and practices based on findings

Name	Non-Discrimination as an Objective
Statement	Set clear non-discrimination objectives and establish predefined acceptance criteria that align with these objectives.
Rationale	By establishing acceptance criteria, non-discrimination can be recognized as a goal rather than merely a financial burden.
Implications	• Define clear objectives and acceptance criteria in advance, including thresholds • Utilize metrics and KPIs to track progress • Assess performance against objectives regularly • Equip decision-makers with the capabilities to effectively achieve these objectives • Continuously review and update criteria to ensure relevance