UNIVERSITAT DE BARCELONA BUSINESS SCHOOL

MSc International Business

EXAMINING ALGORITHMIC BIAS IN AI-POWERED CREDIT SCORING: IMPLICATIONS FOR STAKEHOLDERS AND PUBLIC PERCEPTION IN AN EU COUNTRY

Student: Juris Antonevics

Master's Thesis tutor: Guillem Riambau Armet

Date: 30.06.2023.

ABSTRACT

The presence of artificial intelligence (AI) in financial markets is becoming increasingly common, with AI algorithms used in fields as alternative credit scoring to render complex multifactorial decisions. However, concerns have been raised by agencies and researchers regarding the objectivity of AI based algorithms and their potential to perpetuate systematic inequalities among vulnerable populations. Through a simulation exercise using non-financial data, testing 126 application profiles, this research investigates the impact of AI in credit scoring, examining the presence of algorithmic bias and its implications for stakeholders. Additionally, it explores the perceptions of AI systems among the general population in an EU country surveying 144 individuals. Results from the simulation show presence of unequal treatment towards women applicants, making them less likely to get approved for a financial instrument compared to men. Furthermore, this study also reveals scepticism of general population towards automated decision-making systems and highlights their concerns about data privacy when interacting with AI systems.

Keywords: Artificial intelligence, credit scoring, algorithmic bias, trust, non-financial data

TABLE OF CONTENTS

1. INTRODUCTION

Algorithms endowed with artificial intelligence (AI) are on the rise in financial markets, used for automating underwriting, credit scoring, and investment management tasks. The integration of AI technologies has not only enhanced the efficiency of processes within financial institutions, increased personalisation of their services, and overall contributed to their competitiveness (Xie, 2019; OECD, 2021), but also contributed to improving the capabilities of existing tools, such as credit scoring algorithms, by allowing the incorporation of non-financial data, thus creating opportunities to extend financial inclusion to previously unbanked populations, as noted by the OECD (2021). However, for successful implementation and supporting financial institution processes, trust in AI among users and stakeholders plays a pivotal role (Gsenger & Strle, 2021). Concerns have been raised by the European Union Agency for Fundamental Rights, or FRA (2022), and researchers from Women's World Banking Kelly & Mirpourian (2021) regarding the objectivity of the results some of the algorithms produce, besides emphasising a gap in understanding of AI violations of fundamental rights and a lack of evidence-based assessments in this field. Recent evidence shows that AI algorithms present a significant risk of having hidden biases resulting in unfair treatment towards certain vulnerable populations, namely minorities and women (Kelly & Mirpourian, 2021). Despite the concerns raised, AI algorithms continue to be implemented in tools as credit scoring, which in a way is paving new and improved methods of assessing credit worthiness and thus promoting accessibility of financial services, but on the contrary, they present not yet fully understood set of risks for unfair treatment and perpetuation of systemic inequality.

1.1. Research objective

This master's thesis research aims to examine the effects of use of the AI for financial services, particularly evaluating the possibility of algorithmic bias presence in AI algorithms for credit scoring and determining its consequences for stakeholders. The main objective is to examine how the use of non-financial data in AI-based credit scoring algorithms addresses or perpetuates biases. To make sure the research is well placed and contextualised, the use of customer-oriented AI systems in the financial sector, particularly addressing the aspect of financial inclusion will be explored. The final objective is to investigate the perceptions of AI, trust in its decisions, and attitudes towards sharing non-financial personal data among the general population in an EU country. In this paper, a key research question is to assess to what extent AI models may cause bias in the financial service sector, particularly when addressing customers.

1.2. Applied methodology

This master's thesis research involves three main components: Firstly, a comprehensive literature review on the principles of AI technologies, trust, the origins of algorithmic bias, and the utilisation of machine learning algorithms in credit scoring. Secondly, a quantitative analysis – a survey, gathering the attitudes of 144 respondents from an EU country towards AI systems and the risks associated with algorithmic bias. Lastly, the research includes a simulation study of the AI-based machine learning (ML) algorithm of Bhatnagar (2023), testing 126 credit application profiles for algorithmic bias. The approval prediction-based credit scoring algorithm is trained on a dataset of 438,557 application profiles containing socio-demographic information and credit records. The evaluation of algorithmic fairness utilises statistical and similarity-based measures. The obtained data from the survey was analysed with JASP statistical software using descriptive statistics, frequency analysis, and regression analysis. For a more comprehensive understanding of the applied methodology, please see Chapter 3, where further details can be found.

1.3. Structure

In order to successfully examine the effects of AI technologies in the financial sector, addressing the risks of algorithmic bias and its impacts on stakeholders, the thesis is divided into six chapters. The first chapter introduces the topic of this thesis, briefly describing the background of the study, its topicality, as well as the objective and applied methodology of the study. It also describes the structure and sustainable development goals addressed in the thesis. The second chapter, a literature review, is divided into two parts. The first part covers an extensive analysis of the theoretical background in AI used in the financial sector, particularly its applications, financial inclusion, trust, and algorithmic bias. The second part focuses on machine learning algorithms in credit scoring, explaining the principles of it and methods utilised. The third chapter describes the methodology of the master's thesis research, explaining the scientific approaches used in performing the credit approval prediction simulation study and conducting the survey. The following chapter of results includes data analysis obtained from the simulation study and the survey and outlines the main findings. The fifth chapter focuses on the interpretation of the results and considers the limitations of the research. And the final sixth chapter summarises the main findings and offers recommendations for future research.

1.4. Sustainable development goals addressed

This master's thesis topic covers United Nations Sustainable Development Goal 10, which aims to reduce inequalities in society. Corresponding targets that are being contributed to are: 10.2 – fostering social, economic, and political inclusion regardless of age, race, origin, economic conditions, or other factors; 10.3 – providing equal opportunities and diminishing result disparities; and 10.5 – enhancing the control and supervision of international financial markets and organisations (United Nations, 2023).

2. LITERATURE REVIEW

2.1. Use of AI in financial sector: applications, inclusion, trust, and bias

The last two decades have marked a rapid development of Artificial intelligence (AI) technologies and algorithms, which has led to their increased presence in the everyday lives of the general population as well as widespread adaptation in the fields of healthcare, law enforcement, social media platforms (Gsenger & Strle, 2021), and most importantly, the field of finance (Xie, 2019). Artificial intelligence powered algorithmic decision-making is used for diverse functions in the financial area, involving tasks such as determining loans and insurance premiums, calculating credit scores (Gsenger & Strle, 2021), and others. Because the field of finance is predominantly of a numerical nature and the most relevant AI models being used are based on machine learning method (Aziz, Dowling, Hammami, & Piepenbrink, 2022), this chapter will mostly focus on the utilisation of machine learning algorithms. Additionally, it will address crucial aspects such as financial inclusion, trust, and algorithmic bias in relation to AI systems used in regulated financial institutions.

2.1.1. Principles of AI powered technologies and their application in financial services

The idea that all aspects of intelligence, including learning, can be precisely described in a way that machines can reproduce them traces its roots back to 1956, when it was first presented by J. McCarthy, creating the field of artificial intelligence (Dixon et al., 2020). Since its introduction and rapid development in the past decades, it has evolved into a transformative force affecting global society and becoming one of the fastest growing industries in the world (Woodward, 2023).

The term AI refers to a wide range of techniques that involve machines showcasing intelligence (Aziz, Dowling, Hammami, & Piepenbrink, 2022). One of the core principles of AI is machine learning (ML), which, as a term introduced in 1959 by A. Samuel, characterises pattern recognition tasks by artificial intelligence systems. Machine learning systems are characterised by classification and forecasting executed through a learning process (Gogas & Papadimitriou, 2021), without being programmed to reach a certain result (El Naqa & Murphy, 2015). Although ML as a field is related to AI and its principles can be found in AI, in some instances, if it is operating as an independent learning system, it exceeds the definition of AI and can be considered a separate field. Despite that, in the real world, both terms are mostly used interchangeably (Gogas & Papadimitriou, 2021). ML algorithms fall into three categories: two major categories of supervised and unsupervised learning, which are further divided into multiple techniques under each of the branches, and semi-supervised learning. ML techniques utilised in finance are Bayesian network algorithms, logistic regressions, and support vector machines, which are supervised learning methods (Huang & Yen, 2019). Section 2.2.2 further explores supervised learning, particularly as utilised in credit scoring.

The first application of machine learning techniques for solving economic and financial issues dates back to 1974 (Gogas & Papadimitriou, 2021). Since then, four categories of applications for machine learning algorithms in the financial field have emerged. These categories include applications that are customer-oriented, management-level, financial market transactions and portfolio management, and financial regulations (Xie, 2019). One notable field of ML applications is at the customer level, where algorithms are employed for creditworthiness assessments, insurance services, and chat bot operations (Xie, 2019). Machine learning technique's ability to process vast amounts of predictive information has also facilitated its use for performing risk forecasting, which has historically been a core focus of ML in finance. These characteristic features have been used in real-world credit risk models for lending decisions, thus improving the risk management of financial institutions (Aziz et al., 2022). Apart from forecasting benefits, the main drivers for implementation are the ability of such techniques to bring fairer and more impartial decision-making due to emotionless judgements and a larger degree of transparency, in addition to workload reduction benefits for employees and financial savings on internal and external processes for the company (Gsenger & Strle, 2021; Xie, 2019). From the perspective of consumers, the adoption of AI provides higher-quality services and paves the way for increased financial inclusion for clients with insufficient credit histories (OECD, 2021).

2.1.2. Opportunities for broadening financial inclusion

Financial inclusion is understood as the extent to which individuals have access to banking or financial services (Mhlanga, 2020). However, the World Bank defines financial inclusion even more broadly, referring to it as the availability of convenient and fairly priced financial products and services for individuals and businesses in a responsible and sustainable way (The World Bank, 2023).

When examining the issue of financial inclusion, it is essential to view two interconnected aspects: firstly, the lack of bank account ownership, which is essential for accessing any form of financial instrument, and establishing credit history; and secondly, insufficient credit histories, which can result from the absence of bank accounts and subsequently restrict clients from obtaining products provided by financial institutions.

According to the data from the World Bank (2021), a considerable proportion of global population – about 24%, or almost one in every four adults – remains excluded from having an account at a bank or any regulated financial institution. Most of them are women, individuals from poor households, people from rural areas and the unemployed. The Southeast Asia and Pacific regions present the highest numbers of adults with no bank account, with countries like India, Pakistan, and Bangladesh in South Asia and China, Indonesia, the Philippines, and Myanmar in East Asia and the Pacific leading the absolute number of such individuals (The World Bank, 2021a). According to the article published in the World Economic Forum, over six in ten Southeast Asians are unbanked (Lim, 2022). Especially severe cases are encountered in Pakistan and Cambodia, where the percentage of unbanked individuals makes up 79% and 67%, respectively (The World Bank, 2021b).

Although not as severe as in Southeast Asia, the issue of citizens without a bank account also persists in the European Union. Bank account ownership varies a lot between the countries, with most European Union countries having bank account ownership above 95% and even close to 100% in such countries as Estonia, France, and Finland. Nevertheless, when looking at the whole EU perspective, multiple countries such as Portugal, Croatia, or the Slovak Republic represent less than 95% ownership of bank accounts. Especially low bank account ownership is found in Bulgaria, where it comprises 84%, and an even lower number in Romania, at 69%, which is comparable to the bank account ownership results in some Southeast Asian countries.

Undoubtedly, the issue of not owning bank accounts is complex and consists of multiple factors, including but not limited to insufficient financial resources, a large geographical distance to financial institutions, and a lack of information. Nevertheless, technological advancements, such as emerging new identification systems based on AI technologies and an increase in mobile phone ownership, have the potential to facilitate account ownership (The World Bank, 2021a). Consequently, these advancements can contribute to obtaining more financial information on individuals.

Limited ownership of bank accounts, especially among the population of Southeast Asia, is an undoubtedly significant issue contributing to financial exclusion; nonetheless, Lim (2022), Simumba, Okami, Kodaka, Kohtake (2018), and Aggarwal (2018) underline insufficient credit history as yet another limitation that restricts access to formal financial tools even for those who own an account. Due to a lack of credit history, banks and other financial institutions are unable to evaluate the creditworthiness of potential borrowers. However, Simumba et al. (2018) claim that employing machine learning algorithms that use non-financial data to evaluate the credit worthiness (performing credit scoring) of potential borrowers, drawing on data from mobile devices and social media, has the potential to create financial inclusion for individuals who do not have sufficient credit histories. A similar view is held by Packin (2018), Djeundje, Crook, Calabrese, & Hamid (2021), and Purda & Ying (2022), who also emphasise the potential of non-financial data not only for broadening financial inclusion but also for allowing banks and other financial institutions to increase the accuracy of traditional credit scoring systems. Section 2.2 further explores alternative credit scoring principles.

Looking at smartphone ownership as a potential data source for alternative credit scoring evaluation and the first step for obtaining a bank account, according to the data from GSMA (2021), 68% of the Asia-Pacific population owns a smartphone. Previously addressed countries of East Asia like Indonesia, the Philippines, and Myanmar present higher smartphone ownership levels, reaching levels up to 78%, expected to increase to 85% in 2025 (GSMA, 2021).

Looking at smartphone ownership in Europe, as of 2023, 84% of Europeans owned a smartphone, with expected growth up to 86% in 2025 (Statista, 2023a). Overall Europe as a region presents second highest smartphone ownership rates in the world, after North America (GSMA, 2022). Concerning the case of Romania and Bulgaria, two countries with the lowest bank account ownership, in Romania in 2023, an estimated 80% of the population owned a smartphone, with a projection to reach 84% in 2025 (Statista, 2023b). Meanwhile, in Bulgaria in 2023, smartphones were owned by 70% of the population, which is expected to increase to 73% in 2025 (Statista, 2023c).

Widespread use of smartphones in EU and Asia Pacific region countries with low bank account ownership not only presents an opportunity for expending account ownership through mobile banking and new identification systems – since in countries such as Romania in the EU or Cambodia and Myanmar in Southeast Asia, the number of people owning a smartphone exceeds the number of people owning a bank account. Also, the wide adoption of smartphones provides an extensive data source that can be utilised for AI based alternative credit scoring algorithms. This presents an opportunity to overcome the challenge of limited credit history and paves the way for a higher degree of financial inclusion. Further, the topic of alternative credit scoring algorithms and their application is discussed in Chapter 2.2.

2.1.3. Trust in AI algorithms

Having trust in the algorithm is essential for successfully supporting human decision-making. Such trust is formed if the algorithm presents qualities of being reliable, useful, and consistent (Gsenger & Strle, 2021). Gsenger & Strle (2021) identify that in scenarios where mechanical tasks are preformed, the trustworthiness of algorithms in decision-making is generally perceived as equally reliable as decisions made by humans. However, from the operator's perspective, some contradictions appear. Because of algorithm proneness to mistakes, a complete reliance on algorithms can be considered equally counterproductive to the state of not trusting them at all (Gsenger & Strle, 2021), proving a necessity for some kind of control mechanism implementation, further discussed in Section 2.1.4. Furthermore, a research study on attitudes towards AI in Latvia has reveiled a correlation between trust in AI and positive attitudes regarding such technologies. Interestingly, trust in this sense is not connected with knowledge and rational considerations, which would relate to the principles of AI operations, but rather abstract and socially based factors (Vasiljeva, Kreituss, & Lulle, 2021).

Looking at attitudes towards AI algorithms from the perspective of general populations, the research conducted by Neudert, Knuutila & Howard (2020) of the Oxford Commission on AI & Good Governance has looked into global perceptions of AI in different regions of the world. Their study on Europe reveals that the perception of AI in decision-making holds scepticism among the population, with 43 percent of people finding AI algorithms potentially harmful, while 38 percent believe in the positive utilisation of such technologies. In comparison to other regions of the world, Europeans present a relatively low level of trust in AI. For instance, East Asia has one of the highest trust levels, with 59 percent of the population believing in the benefits of AI in decision-making (Neudert, Knuutila, & Howard, 2020).

Examining the general attitudes towards AI utilisation in Latvia, as studied by Vasiljeva, Kreituss, & Lulle (2021), reveals a notably higher positive perception rate of AI compared to the European average, with 53 percent of surveyed Latvians revealing a positive or a very positive attitude towards AI. Importantly, despite previously held unfavourable views on AI, the study highlights a recent shift towards much more favourable views, as affirmed by AI experts and management representatives interviewed in the study (Vasiljeva et al., 2021). These historically unfavourable views towards AI can be partially observed in the study of the Baltic International Bank (2018), based on data from 2017, which shows significantly lower support, with only 34% of the population holding favourable views towards the development of AI technologies. Remarkably, just 10 percent of the population would prefer an algorithm-made financial decision (involving investments in stocks or pension funds) over a decision made by a human consultant (Baltic International Bank, 2018), thus revealing excessive scepticism for decisions made by algorithms.

In the context of AI-based automated decision-making systems, Gsenger & Strle (2021) identified the significance of individual aspects such as cultural background, community perception, and mass media influences in shaping expectations surrounding these systems. Moreover, they investigated factors contributing to the positive perception of automated decision-making, emphasising the importance of emotionless and value-neutral judgements as favourably viewed attributes of such systems. Another study conducted by Smith (2018) on attitudes towards algorithmic decision-making in the US in 2017 founds that the majority of the population has concerns about their data privacy when interacting with automated finance score systems. Moreover, people raise concerns about whether the data representing a person will lead to accurate results.

The study of Vasiljeva, Kreituss & Lulle (2021) also explored the perspectives of employees of various industries. The findings demonstrate that organisations with no immediate plans for AI technology adoption have comparatively low levels of positive attitudes towards AI. In contrast, companies that have already adopted AI have much higher percentages of favourable or highly favourable views among their employees, accounting for 87 percent of cases (Vasiljeva et al., 2021). The finance industry, together with the legal and business service industries, is found to be the most positive regarding AI, with an especially positive image in international companies. Altogether, companies that are larger in size tend to have employees with more favourable views towards AI compared to small and medium enterprises (Vasiljeva et al., 2021).

2.1.4. Concept of bias in AI algorithms: origins, explainability, and mitigation

Outcomes generated by algorithms can be biased, just like ones made by humans (Kelly & Mirpourian, 2021; Pethig & Kroenung, 2022). Bias in the context of AI algorithms is defined as systematic errors or inaccuracies in the decision-making process that come from flawed assumptions, incomplete data, or poor design of the algorithm, which inherits prejudgements (Kelly & Mirpourian, 2021) and thus produces unfair or discriminatory results. These results pose undesirable implications for the exclusion of vulnerable populations from credit markets, the perpetuation of existing inequalities, and even, to a certain extent, limiting economic growth (Kelly & Mirpourian, 2021). Nevertheless, the potential for AI algorithms to amplify biases and discrimination, or diametrically opposite, eliminate unfair decisions and discriminatory exclusion depends on the way they are implemented (OECD, 2021). However, it is worth noting that there are currently no universal standards for measuring the fairness of AI systems. This leaves each organisation on their own to determine the extent of fairness and identify the types of biases most probable in their systems (PWC, 2023).

Addressing bias in AI algorithms two of the most common origins of biased results produced by AI algorithms can be identified. Firstly, bias because of the structure of the algorithm itself and the way it is written; and secondly, bias due to incomplete, misbalanced, or potentially prejudice data entries (Kelly & Mirpourian, 2021; PWC, 2023). However, research by the PWC (2023) consulting company addresses the third origin as the interpretation of the results produced by an algorithm. Further developing the origins of bias, Kelly & Mirpourian (2021) and the OECD (2021) highlight three types of possible algorithmic bias: sampling bias, labelling bias, and outcome bias.

Sampling bias represents the poor quality of invalid, incomplete, or in any other sense flawed data that can lead to incorrect or biased results from AI algorithms (OECD, 2021). Poor-quality data can take the form of overrepresentation or underrepresentation of certain groups of people in the data set (Kelly & Mirpourian, 2021) or a form of data that presents certain historical biases or discriminatory decisions made by humans. Machine learning algorithms that have been trained with such data will present bias results even if later good-quality data is imputed (OECD, 2021).

Labelling bias represents the result of improperly assigning data points to annotate and categorise specific features and traits of a person. Assigning data points is essential for the algorithm to be able to identify such features (Kelly & Mirpourian, 2021). While the process of labelling data presents opportunities to detect errors and biases within the data, it also poses risks of unconsciously introducing new biases because of subjective decision-making. Hence, financial institutions should understand the risks and mitigate them by ensuring the use of various sources of labelled data and labelers of diverse backgrounds (OECD, 2021).

Outcome proxy bias occurrence is related to a not well-defined assignment for the machine learning algorithm in which outcome measures are not directly related to the outcome of interest (Kelly & Mirpourian, 2021). Packin & Lev-Aretz (2018) and OECD (2021) highlight that it might not come as intentional discrimination from the company but rather a result of an algorithm assigning or determining sensitive characteristics such as race or gender based on multiple facially neutral data points such as transaction activities or other variables and afterwards utilising this information in its decision-making, e.g., in the determination of creditworthiness.

Nevertheless, even if a machine learning algorithm is trained on high-quality, well-labelled data and well-defined assignments, it can present unintentional biases that just occur through correlations of sensitive and non-sensitive variables that are problematic to identify in the massive databases. These correlations can perpetuate existing biases that are present in society and just happen to be reflected in the datasets used for training algorithms (OECD, 2021). Research by Packin & Lev-Aretz (2018) has highlighted the increased vulnerability of machine learning algorithms used in decision-making, particularly within consumer credit application predictions, towards various types of biases. Packin & Lev-Aretz (2018) also emphasised that automated decision-making made by machine learning algorithms can be hard to explain. Moreover, such models are incapable of distinguishing causation from correlation (Packin & Lev-Aretz, 2018).

When examining the origins of bias in AI algorithms, as highlighted by PWC (2023), undesirable outcomes can also stem from human interactions with these algorithms. The inclination of humans to disregard contradicting computer-generated results, inadequately analysing them, and considering them as correct is referred to as automation bias (Gsenger & Strle, 2021). It is characterised by excessive reliance on automated results and is part of the cognitive bias group. Specifically, it can be associated with anchoring bias (heavily reliance on the initial information); availability bias (judgement reliance on previous examples recalled from memory); confirmation bias (favouring information that confirms pre-existing beliefs); and representativeness bias (information matching with mental models or stereotypes) (Vered, Livni, Howe, Miller, & Sonenberg, 2023). One of the main tools for mitigating automation bias is the implementation of explainable artificial intelligence systems (Vered et al., 2023).

Explainable artificial intelligence (XAI) is a set of methods used to mitigate the black-box principle or for humans' unintelligible rules made by AI models (Weber, Carl, & Hinz, 2023). It refers to understanding the main drivers of algorithms decisions and the possibility of answering questions about the model's operations. Some degree of explainability for the decision-making algorithms can be required by normative regulations like the General Data Protection Regulation (GDPR), which determines the rights of data subjects to receive insights about the reasoning behind machine-generated decisions (Bussmann, Giudici, Marinelli, & Papenbrock, 2021). The overall field of finance is a highly regulated industry, especially in the European Union and the US, which makes XAI method implementation crucial in this region for any task automation using AI (Weber et al., 2023). According to Weber et al. (2023), there is currently a lack of sufficient research on XAI applications in the field of finance.

Regarding the overall tackling of algorithmic bias, Kelly & Mirpourian (2021) emphasise that artificial intelligence tools are not able to detect and ignore biases or biased data on their own. A strategy regarding the mitigation of bias in financial algorithms has been proposed by Kelly & Mirpourian (2021). They propose assessing the presence of bias in 3 stages: the pre-processing stage, with careful analysis of the data, where the algorithm will be trained; the in-processing stage, where fairness constraints are introduced in the model; and lastly, the post-processing stage, where model output predictions are reviewed, and any possible biases are reduced. Nevertheless, combating bias in algorithms often comes at the cost of the accuracy of the model (Kelly Mirpourian, 2021). Kelly & Mirpourian (2021) highlights that operational processes and norms within organisations play an important role in mitigating bias, emphasising the necessity of involving all the members of the organisation in a transition to AI algorithms. Besides, understanding the market and a skilled team, frequent analysis of the performance of algorithms, and responses to systemic shocks are critical factors to consider in AI algorithms, in addition to investigations of unwanted correlations among the most essential variables and addressing the issue of false negatives category (Kelly Mirpourian, 2021). In the reduction of errors produced by algorithms, human introduction into the loop of the algorithm is proposed by Gsenger & Strle (2021) as a potentially effective way to reduce errors and biases.

2.2. Machine learning algorithms in credit scoring

Credit scoring practises have undergone a significant transformation with the emergence of machine learning algorithms, allowing automation of the processes, and enabling the integration of non-financial data for creditworthiness evaluations, which is commonly referred to as alternative credit scoring (Njuguna & Sowon, 2021). Such developments have provided opportunities for increasing financial inclusion among populations with limited credit histories (Simumba et al., 2018), while also improving accuracy in traditional credit scoring systems (Packin, 2018; Djeundje et al., 2021; and Purda & Ying, 2022). This chapter will primarily focus on the basic principles of credit scoring using financial and non-financial data. Additionally, exploring the machine learning algorithms used in creditworthiness evaluations, with a particular emphasis on the random forest algorithms identified as the most suitable for credit scoring practises (Hindistan et al., 2019).

2.2.1. Credit scoring using financial and non-financial data

Looking at the principles of credit scoring, it is a widespread statistical method of appraisal used by banks and other financial institutions to determine the creditworthiness of a potential borrower based on their financial, sociodemographic, and credit history data and risk factors, is a fairly recent practice (Popovych, 2022; Abdou & Pointon, 2011). Although the earliest records of borrowing and lending date back to 2000 BCE, the first applications of credit scoring appeared roughly six decades ago (Abdou & Pointon, 2011).

Substituting the previous judgmental practises performed by loan analysts with (traditional) credit scoring methods has contributed to significant economies of time, given the degree of automation they provide, and eliminated insignificant factors from credit evaluations that do not correlate with repayment performance, thus eliminating potential belief-based and statistical bias (Popovych, 2022; Abdou & Pointon, 2011). In the views of Popovych (2022), it allowed for credit evaluations to exclude discriminatory age, sex, race, and other sensitive factors. Another significant benefit of the introduction of credit scoring is the elimination of bias resulting from the sole consideration of the repayment histories of accepted, but not all, applications, which would have clearly been beyond the cognitive capacity of a judgmental loan analyst (Abdou & Pointon, 2011). In regard to the elimination of belief-based bias, the findings of Pethig & Kroenung (2022) suggest that potentially vulnerable groups, e.g., women, tend to perceive decisions made by algorithms as more objective in comparison to human assessments.

Looking at the principles of traditional criteria-based evaluations, they rely on the Five Cs analysis as the basis for assessing potential borrowers. Such evaluations involve the collection and analysis of factors such as capacity to repay, capital, collateral, conditions, and character (credit history) (Yhip & Alagheband, 2020). In the past, methods such as the FICO score have been employed for these evaluations (Gsenger & Strle, 2021).

Recently, there has been a surge in attention from the research community towards exploring opportunities for enhancing credit scoring practices. One of such focuses is the integration of alternative data, which promises to foster financial inclusion as previously identified Section 2.1.2.

Credit scoring using alternative data, also known as 'alternative credit scoring', is a credit worthiness evaluation method that, unlike traditional practices, uses alternative, non-financial data requested by companies to run prediction algorithms that determine what Seon (2023) calls "lending trust". These alternative data include non-credit financial information as utility, rental, and telecommunication payment records, and non-credit, non-financial records as web search histories, online shopping trends, information from social media networks, records of individuals' online activities, mobile phone activities, contacts, calendar information, location (Packin & Lev-Aretz, 2018; Njuguna & Sowon, 2021; Kelly & Mirpourian, 2021; Asian Development Bank, 2023), and other data directly or indirectly reflecting the behavioural patterns (Aggarwal, 2018). Historically, some traditional credit scoring methods have considered several non-financial factors, such as the applicant's age, location of residence, occupation, or employment history. Nonetheless, these characteristics had a lesser influence on the results compared to the present models that incorporate alternative data sources (Packin & Lev-Aretz, 2018).

Purda & Ying (2022) claim that the data sources that could be used for alternative credit assessments are virtually limitless. Moreover, experts concur that no single methodology can be suitable for every credit assessment scenario; thus, each and every alternative credit assessment algorithm presents a unique set of variables. This heterogeneity therefore can be attributed to specific applications and different algorithmic models, including expert systems, simple statistical models, and advanced machine learning models (Purda & Ying, 2022).

As outlined in Section 2.1.2 and affirmed by Simumba et al. (2018), Packin (2018), Djeundje et al. (2021), and Purda & Ying (2022), the significance of the use of alternative credit scoring algorithms lies in their opportunities for broadening financial inclusion for those individuals who lack sufficient credit history to receive financial services. One example of an alternative credit scoring algorithm application that solves the issue of insufficient credit history is the model developed by Simumba et al. (2018). The developed model uses alternative data for performing creditworthiness evaluations of smallholders in Cambodia. Their model evaluates a range of factors divided into three categories: predictors of fraud (46 factors), predictors of interaction (20 factors), and predictors of revenue (18 factors). These factors include personal information about the potential borrower, such as age, sex, and the number of family members; information about the farm, such as its location, area, sown area, crop type, and harvest method; data about interactions with the specific mobile app for monitoring and independent moderators, such as frequency, number of reports, days since joining the app, and other factors (Simumba et al., 2018).

Overall, the factors of the above-mentioned model are similar in scope to the variables identified by Purda & Ying (2022) when seeking to exemplify generic alternative data.

However, some of the factors introduced by Simumba et al. (2018) and Purda & Ying (2022) contradict Popovych's (2022) views on the essential principles of credit scoring algorithms, in particular aspects of excluding discriminatory age, sex, race, and other sensitive factors, which consequently secures an objective credit score for potential borrowers. It should be stressed, however, that Popovych (2022) speaks of these principles in relation to traditional credit scoring algorithms, which do not consider alternative data; nevertheless, we can assume that the basic principles of credit scoring, especially the exclusion of potentially discriminatory factors, apply to any credit scoring algorithm regardless of the type of data (financial or non-financial) used.

While only limited number of studies have focused on the use of non-financial data in machine learning credit scoring systems, mostly due to lack of access to such data on individuals (Djeundje et al., 2021), there are some studies as those of Simumba et al. (2018), Khemakhem & Boujelbene (2018) and Djeundje et al. (2021), exploring it from the perspective of predictive accuracy improvement. Therefore, only few studies, as those conducted by Packin & Lev-Aretz (2018), Gsenger & Strle (2021), and most importantly, Kelly & Mirpourian (2021), have specifically addressed the utilisation of non-financial data in these systems and its relationship with algorithmic bias in English scientific sources. It is worth noting that no papers were found that performed algorithmic bias testing on alternative credit scoring models, particularly involving algorithmic audit principles.

2.2.2. Machine learning methods utilised in credit scoring algorithms

Addressing the machine learning methods used in credit scoring, they can be characterised as a more advanced approach to credit scoring practices, substituting previously used logistic regression and discriminant analysis, which are part of traditional statistics (Aji & Dhini, 2019).

The essence of machine learning models involves the utilisation of big data for autonomously learning and making predictions and decisions in a way that is not in any way defined by a programmer (OECD, 2021). As highlighted in Section 2.1.1, algorithms used in finance and consequently also in credit scoring practices mostly utilise supervised learning models. Supervised learning takes place when each training example of input data, e.g., colour, shape, or weight, is matched with a corresponding label that classifies it, e.g., apple and banana. In this way , the algorithm learns to distinguish different classes by recognising the key qualities of the object and differentiate between different classes (El Naqa & Murphy, 2015). Particular benefits of machine learning algorithms over previous practices are their change-friendly nature, which provides opportunities for adjusting training data – eliminating certain features or providing new features – and finding new relationships between the data (Packin & Lev-Aretz, 2018).

Looking particularly at the machine learning methods used for credit scoring purposes, various methods for classification tasks are used. Some of the more widely accepted methods used are logistic regression and random forest, which belongs to the supervised learning models (Hindistan et al., 2019). The research conducted by Khemakhem & Boujelbene (2018) using financial and non-financial data for scoring purposes has identified decision trees and neural networks as particularly suitable for credit scoring. In their research between these two models, they identified decisions made by decision trees as superior in accuracy for predictions in comparison to neural networks. Also finding high sensitivity of the models to imbalanced data. However, similar research by Hindistan et al. (2019) presents different findings, identifying logistic regression and random forest as superior to decision trees. Nonetheless, Hindistan et al. (2019) highlight that different methods apply for different scenarios, and none of the methods overall can be characterised as superior to others.

Looking specifically at the random forest principles, it can be considered an improved method of decision trees, which utilises the same methodology combining multiple tree predictors, where each tree relies on the values of a randomly sampled vector for its predictions (Biau & Scornet, 2016) (see figure 1).

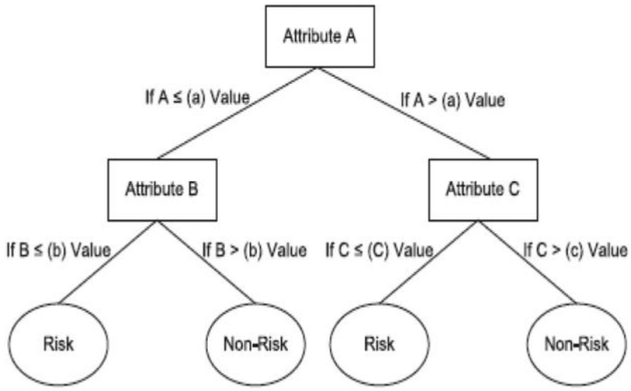


Figure 1. Decision tree visualisation.
Source: Khemakhem & Boujelbene (2018)

Each tree of the model predicts a certain outcome, and in the case of credit scoring, which is a classification task, the final prediction is a result of the majority of "voting" between the trees (Biau & Scornet, 2016). Such a model has been observed to reduce biases in comparison to previous machine learning models (Breiman, 2001) and has been proven to be superior in comparison to other classifiers (Uddin, Chi, Al Janabi, & Habib, 2022). Moreover, its usage in credit risk prediction provides favourable new options through the usage of continuous and categorical variables of features, and in terms of its architecture, it presents a rather simple structure (Uddin et al., 2022).

When looking at machine learning algorithms used for credit scoring and related areas, it is important to consider their limitations, which are mainly associated with their inability to independently determine the sufficiency of the tested data for making sound judgements (Packin & Lev-Aretz, 2018). While also being prone to various types of hard-to-explain intentional and unintentional biases identified by Kelly & Mirpourian (2021), the OECD (2021), Packin & Lev-Aretz (2018), and other researchers in Section 2.1.4, which proves the necessity for such algorithm testing.

3. METHODOLOGY

3.1. Research objective

The research aims to examine the effects of the use of AI for financial services by evaluating the possibility of algorithmic bias in AI algorithms for credit scoring and determining its consequences for stakeholders. Additionally, the research seeks to understand the perceptions of AI, trust levels in its decisions, and attitudes towards alternative credit worthiness evaluation and the steps it involves. To achieve these objectives, I will first examine how the use of non-financial data in AI credit scoring algorithms addresses or perpetuates biases. And further, investigate the perceptions of AI and alternative credit scoring among the general population of an EU country.

3.2. Literature review

To understand the context of the research and set a theoretical framework, I conducted an extensive literature analysis. With primary fields of focus on artificial intelligence & machine learning, and business finance. From the selected papers, the earliest paper dates back to 2001; however, the absolute majority of papers utilised in the review were published between 2018 and 2023, including the most up-to-date literature on the topic.

3.3. Research design

The paper utilises a quantitative research approach with the use of a survey and simulation, which are independent from each other.

The questionnaire was designed to collect data on the perceptions of AI, alternative credit scoring, and the use of various types and sources of information by financial institutions among the general population of Latvia. It consists of 21 questions divided into four sections (see Appendix 1). The first section focused on the socio-demographic profile of the respondent and their affiliation with the potentially vulnerable groups identified by Kelly & Mirpourian (2021) and the European Union Agency for Fundamental Rights or FRA (2022). In the next section, data was obtained on general attitudes towards AI and respondent affiliation to techno optimists or techno pessimists based on the questions used by Vasiljeva et al. (2021) and Baltic International Bank (2018). The third section concentrated on respondent interaction with credit institutions and the evaluation of an alternative credit scoring scenario. In the fourth section, comfort levels of respondents when sharing sensitive information were evaluated, as were their concerns and their proneness to choose alternative credit scoring algorithms under certain conditions.

The majority of the questions in the survey utilised a Likert scale, ranging from 1 to 5. For questions and topics where the Likert scale was applicable, multiple-choice questions were used instead. Multiple-choice questions were made with the option to provide a personalised response if none of the possible answers provided were not found to be suitable. Regarding personal attitudes towards creditworthiness evaluation, which were measured in the final question, it was presented in two versions. These questions were divided between two groups of respondents to explore the impact of the perception of risk for discrimination, views towards AI, and potential personal benefits on the inclination to be evaluated under such a method.

For the simulation part, a model of Bhatnagar (2023) for approval prediction was used, published in the data scientist community Kaggle. The model is written in the high-level, general-purpose programming language Python and is based on random forest type machine learning algorithm. For the training of the model, two datasets of anonymous financial and non-financial data were obtained from a financial institution and provided by Song (2020) on the *Kaggle* platform. The first dataset contains information on application profiles with 438 557 inputs featuring personal ID, age, gender, number of family members, number of children, yearly income, income type, occupation type, employment duration, education level, family status, housing type, ownership of car, realty, private phone, work phone, and email. The second dataset contains information on the credit record, which reveals the date when the bank account was opened and actual repayment performance. The data used for training the model is up-to-date; according to Bhatnagar (2023), it was extracted on January 1, 2020. The data presents the problem of imbalance identified by Bhatnagar (2023). The model has passed 10-fold validation.

3.4. Data collection

Data for the surveys was collected using Google Forms services. The questionnaire was initially created and shared in Latvian and was later translated into English to include it in the thesis paper (see appendix 1). The data collection period spanned from April 5th to June 5th, 2023. The survey was primarily distributed through online channels such as social media pages like LinkedIn and Facebook and email. A total of 144 responses were received, forming the basis for the survey results.

The simulation was performed on the Google Collaboratory platform. The testing for bias was inspired by the research of Zhang & Kuhn (2022) on algorithmic bias in job recommender systems. My testing involved the creation of a total 126 profiles with various characteristics according to the application record entries, which were compiled in MS Excel spreadsheets. Testing of the model primarily concentrated on the non-financial dimensions of the profiles, focusing on factors such as age, gender, number of children, number of family members, family status, education level, job title, ownership of cars and realty, ownership of phone and email. Details of each profile were manually entered in the algorithm through a numerical code, uploaded, and run by the algorithm, determining a prediction label (approved or denied) and prediction score (from 0.50 to 1.00), which indicates the confidence of the result. The study of the algorithm was divided into 2 phases: the initial phase of randomised testing of 46 profiles, followed by structured testing of 80 different profiles divided into 8 testing groups based on factors of investigation (see Appendix 2).

3.5. Statistical data analysis

The data collected through Google Forms was compiled and organised in MS Excel spreadsheets, where it was cleaned and prepared for statistical analysis. As a part of this process, some responses were transformed from textual format into numerical values to facilitate analysis. Statistical analysis was performed using the open-source analysis programme JASP. For the analysis of survey data variables, both independent variables and the comparison between multiple variables were used for descriptive statistics, frequency analysis and regression analysis, using tools such as contingency tables with Chi-Squared tests to determine statistical significance and Pearson's correlations. Besides, for all of the variables, where applicable, were used statistical measures such as count, mean, mode, median, and standard deviation. Most noteworthy findings have been collected in tables, which are provided in the appendix for reference (see appendices 3 to 21) and further examination.

The data collected from simulation was compiled into MS Excel spreadsheets and analysed either manually or with simple Excel tools such as formulas, quick analysis, or conditional formatting. The evaluation of algorithms fairness involved measures identified by Kelly & Mirpourian (2021), mainly the similarity-based and statistical measures.

3.6. Methodological limitations

The study has several methodological limitations that should be acknowledged. Firstly, in terms of the survey, the sampling method and size may not be enough for generalisation of results. The relatively small sample size introduces the possibility of a large margin of error. Moreover, the survey was specifically targeted at particular age groups and limited to a few geographical locations in a high-income country. As a result, when generalising the findings to other populations, caution must be exercised. Additionally, it is important to note that the survey was shared through social media channels and was primarily answered by individuals within the author's social circle. This aspect raises the potential for selection bias and the possibility that results may not be entirely representative of the wider population.
In terms of the simulation, the study of the approval prediction model involves the analysis of a simplified model that incorporates only a few financial parameters and non-financial parameters that are limited to demographic information and property ownership, thus lacking the wider range of parameters considered in real-world financial institution algorithms. The analysis of the algorithm focused on the detection of bias only in its outputs and cannot be considered an algorithmic audit. Besides, the data used in the simulation is derived from an unspecified geographical location, which poses challenges in terms of interpreting the results and understanding the context and background of the data. Because of the lack of contextual information, the findings may be limited in their capacity to be generalised to a real-world setting.

4. RESULTS

4.1. The simulation

The aim of this testing is to determine how the use of non-financial data in AI credit scoring algorithms addresses or perpetuates biases through testing the presence of bias in Bhatnagar's (2023) approval prediction algorithm published in the data scientist community Kaggle.

From this simulation, I expect to find a limited occurrence of biased outcomes. Primarily, I expect to identify some degree of bias in features associated with the age, family status, and education level of individuals and a large degree of disparities between male and female applicants, considering female applicant vulnerability towards bias identified by Kelly & Mirpourian (2021).

First, it is essential to look at the overall performance of the model. It has passed 10-fold validation and presents the following performance indicators measured by accuracy, area under the curve (AUC), recall, precision, and F1 metrics (see table 1).

Table 1. Performance indicators of the random forest model for approval prediction.

Measure	Accuracy	AUC	Recall	Precision	F1
Result	0.9572	0.6198	0.1538	0.8000	0.2581

Source: Bhatnagar (2023)

The Accuracy of the model presents a percentage of 95.7% of outcomes correctly predicted in the dataset, which indicates a high level of correctness in the performance of the algorithm in terms of true positive or true negative results (Google, 2022a). AUC, or *Area Under the Curve,* of 0.620 indicates a moderate level of models' ability to distinguish between positive and negative instances (both true and false), indicating some level of discriminatory power (Google, 2022b). In terms of models' ability to correctly identify positive instances measured by recall, the score of 0.154 indicates that the model has an increased number of false negative results. A precision of 0.80 reflects a high proportion of correctly predicted true positive instances, indicating low numbers of false positive results. Finally, the harmonic mean or balance between precision and recall measured by F1, scoring 0.258, suggests that the model's overall performance in classification (mainly recall) could be further improved (Google, 2022c). Considering the performance indicators of the algorithm, particularly the recall metric, I foresee a potentially high number of rejected profiles.

Comparing the prediction model of Bhatnagar (2023) with the random forest models of Aji & Dhini (2019) and Malekipirbazari & Aksakalli (2015), which represent 0.73 and 0.78 accuracy, respectively, the model of Bhatnagar (2023) can be considered superior in this dimension. However, in terms of AUC, the model of Bhatnagar (2023) performs worse than the models of Aji & Dhini (2019) and Malekipirbazari & Aksakalli (2015), which present 0.80 and 0.71, respectively. Regarding Recall, Precision, and F1, they were not used in the assessment of models by other authors. Additionally, it is worth noting that models developed by the other authors were trained using different data and use a different set of variables for prediction, therefore, this comparison of performance of various random forest models serves only as an informative reference.

Looking at the testing itself, a total of 126 profiles were tested, with 33% of them receiving positive predictions and 67% receiving negative predictions (see appendix 21). The majority of the results of algorithmic testing for the lending trust came with low confidence levels, ranging from 50 to 70 percent (see appendix 20).

Analysing the importance of age in the algorithm, it was determined that age does not significantly influence predictions, neither in a positive nor negative direction (see appendix 3), except in cases where age is associated with certain occupation statuses, such as being a pensioner. The algorithm tends to favour retired people over working people of various age groups and occupations (see appendix 4). During the initial testing phase, the abovementioned pattern was revealed when comparing two pairs of profiles with identical age, income, education, and other relevant characteristics but differing solely in their occupation status: one of them working and the other one being retired. Results of the prediction showed that the person who was only working got rejected while a retired person was approved (see appendix 4). It is important to mention that both the approval and denial came with a low confidence level of just 53% for the approved profile and an average of 58% for the denied profile (see appendix 4). Moreover, even when contrasting the profile of a retired individual with that of a young and employed individual with an average prestige job and the same income level, the former still meets rejection (see appendix 5).

Next, testing the gender parameter reveals the presence of gender discrimination in the majority of results. When evaluating two or more profiles that are identical in all aspects except gender, it was observed that women constantly receive lower confidence scores on their predictions (see appendices 3 and 6). In multiple scenarios, it goes as far as applications from women getting denied while applications with the same parameters are approved for men (see appendices 3, 6, 7, and 8). It is noteworthy that even when a profile of a man was approved, it was typically accompanied by a low confidence score, which in a real-world scenario would indicate the potential unreliability of the profile. Nevertheless, the fact that one gender was formally accepted by the algorithm provides an unfair advantage over the other. These results are consistent with the findings of Kelly & Mirpourian (2021) regarding female discrimination by AI algorithms.

Furthermore, the examination of family status importance has revealed the existence of unequal treatment towards different groups by the algorithm. The impact of family status varied between men and women. For women, the most favourable family status was found to be married or widowed, as it slightly increased the chances of receiving favourable predictions from the algorithm in comparison to other family statuses (see appendix 6). Similarly, for men, family status of married and widowed was preferred, and in addition, status of being in a civil marriage appeared to influence the results positively, which was not noticed in regard to women (see appendix 6). On the contrary, the family status of being single was associated with the lowest approval rates among all other groups, being relevant for both men and women (see appendix

6). Results of the initial phase of testing showed that profiles with the status as single were constantly denied for both genders (see appendix 9). Overall, married individuals, regardless of gender, were shown to be more likely to receive a positive prediction. Notably, as identified before in the testing for gender discrimination, the algorithm demonstrated preference for men, in particular married men, over married women in terms of the overall approval rates.

When looking at the factor of number of the children among various profiles, the testing reveals different impacts for men and women. For women, having at least one child positively influences the likelihood of receiving approval compared to women who do not have any kids (see appendices 10 and 11). Furthermore, the more kids a woman has, the higher her chances of approval. For men, different tendencies have been observed. Among men, the most preferred profiles are ones without children, followed by profiles with four children (see appendix 10).

Further testing of the algorithm through the scope of various job positions highlighted a bias in favour of job titles that are associated with medium and high prestige occupations, while demonstrating less favourability towards occupations of lower prestige for both men and women (see appendix 12). However, it was observed that profiles of women present a higher disparity of impact between high and low prestige job categories compared to men. Overall, high-prestige jobs, as highly skilled IT professionals were found to be the most favourable out of other categories, both for men and women (see appendix 12).

Among the three occupation statuses examined, it was found that retired individuals tend to receive more favourable predictions compared to those who are working or are unemployed (see annex 4 and 14). On the other hand, profiles of individuals who are currently unemployed were the least favoured by the algorithm. It was also found that unemployment has a more significant negative effect on men than women (see appendix 14). This result is one of the few where, although both profiles are denied, the female profile gets significantly better chances of receiving approval than the male applicant, considering lower confidence score on her denial. Interestingly, the duration of employment was not found to significantly influence the predictions of the algorithm. Nevertheless, even between various employment histories, the profiles of men had more favourable results than those of women (see appendix 13).

The findings from the testing of the impact of housing type features on the decisions of the algorithm revealed that living in a house or apartment presents the most favourable type of housing for male and female applicants (see appendices 15 and 16). For women, housing types such as rented apartments, office apartments, co-op apartments, and living with parents were found to have the least favourable features. Different sets of housing types were found to be less favourable for men, in particular living in municipal apartments and co-op apartments. Notably, living with parents had a much more negative impact on the predictions of the algorithm for men compared to women (see appendices 15 and 16).

Through analysis of the influence of asset ownership, particularly cars and property, it was found that ownership of these assets plays a significant role in determining positive predictions on applications for financial instruments. Remarkably, for both men and women, it is equally essential to own a car in order to get an optimal prediction. A car can be owned either alone or together with a property. Even more surprisingly, it was found that the impact of owning a property differs among the genders: while for women it presents a minor advantage, for men it affects the prediction negatively, reducing the confidence levels in approved application scenarios and increasing them when the application is denied (see appendix 17).

Next, the impact of owning a mobile phone and an email was examined. The insights from the testing suggest that owning both a mobile phone and an email has a large positive effect on the predictions of the algorithm for both men and women. Interestingly, while for women owning a phone is considered more important since it delivers higher prediction results, for men, ownership of an email carries greater significance. Generally, individuals, regardless of gender, who do not own a phone, or an email address are more likely to be denied by the algorithm (see appendix 18).

Looking at the different education levels impact on the algorithm's predictions demonstrated that both male and female applications with education levels of lower secondary and higher education were considered the most favourable by the algorithm in a way that increased their likelihood of receiving a positive prediction (see appendix 19). On the other hand, some significant gender disparities were observed regarding application profiles with incomplete higher education. While women with incomplete higher education are considered less favourable by the algorithm, this negative impact does not apply to male applicants. Besides, when examining the least-favoured education levels, it became evident that for both genders, secondary or secondary special education presents the least prospects for a positive prediction (see appendix 19).

Overall, the results reveal the presence of bias in the predictions of the algorithm across various features, including gender, family status, education level, job title, employment status, housing type, number of children, ownership of assets, and communication tools. Notably, no bias was identified in features such as age or employment duration. Overall, the results show significant disparities between male and female applicants.

In terms of the total number of approved and denied profiles, female profiles on average received approval in only 19% of cases, compared to 48% for male profiles, proving the lack of balance for the positive class (see appendix 21). It is important to note that the overall results for all applicants exhibited considerable variations between different testing phases, as on average in the initial phase between 50% and 64% of total applications were approved, in contrast to the final testing phase, where only 19% of total applications were approved (see appendix 20). Surprisingly, the second testing of the initial phase presented equal approval rates between male and female applicants. In contrast, in the first testing of the initial phase and the final testing phase, women's profiles were hugely disadvantaged, receiving approval only in 11% and 5% of cases, respectively (see appendix 20).

Throughout all the testing phases, male profiles consistently received higher confidence scores for approved applications and lower confidence scores for denied applications in comparison to female profiles, proving the presence of conditional use accuracy inequality (see appendix 20). This indicates that even if identical male and female profiles are approved by the algorithm, the algorithm exhibits more confidence in male applicants.

4.2. The survey

The aim of the questionnaire was to investigate the perceptions of AI, trust in its decisions, and attitudes towards sharing non-financial personal data among the general population in Latvia.

A total of 144 responses were obtained from the survey. According to the data found in the Central Statistical Bureau of Latvia (2023), the population of Latvia in the beginning of 2023 composed 1 883 008 people. Using this information and *calulator.net* scientific tool, the margin of error for the survey results was calculated. The results have a 95% confidence level and are within ±8% of the surveyed values (Maple Tech International, 2023).

In terms of the socio-demographic profile of the respondents, the majority of them, 72.2%, were women, while 27.8% were men. In the breakdown by age groups, the largest proportion of respondents, accounting for 43.7% of the total, were between 22 and 25 years old. The second largest age group consisted of respondents aged 42 years or older, representing 22.2% of all respondents. The remaining age groups were relatively evenly distributed, each representing between 4.9% and 9.7% of the respondents. It is worth noting that all of the age groups included in the questionnaire are represented in the survey (see appendix 22).

By education level, the majority of the respondents have obtained a higher education degree, with the absolute majority (44.4%) having a bachelor's degree, 22.9% a master's degree, and 3.5% a PhD. Another significant group are people with secondary or secondary special education, who make up 27.8%. Out of other educational levels, only 0.7% of respondents represents education below secondary (see appendix 22).

Looking at the types of communities represented by the respondents, it is apparent that the majority of them come from Riga and its metropolitan area, accounting for a total of 52.8%. Next, 26.4% of respondents represent other large cities in Latvia, while 12.5% come from small towns or cities. Rural areas of Latvia are represented by 4.2% of the respondents. Interestingly, 4.2% of respondents indicated belonging to another community outside the territory of Latvia (see appendix 22).

When examining the potentially vulnerable groups represented by the respondents, a significant majority of respondents, or 86.1% (124) indicate that they do not belong to any of the vulnerable groups mentioned in the survey. However, according to Kelly & Mirpourian (2021), women are considered one of the potentially vulnerable groups towards algorithmic bias and thus all women, who represent 72.2% of respondents, can be considered a vulnerable group itself. Among those who indicated belonging to any of the potentially vulnerable groups offered, 6.9% stated their affiliation with LGBTQ, 6.3% indicated their identity as an ethnic minority, 2.8% reported them as political activists, and only 0.7% identified as a person with a disability (see appendix 23). It is worth noting that none of the respondents related them to the status of a refugee.

To gain insight into the dominant attitudes of the respondents concerning information technologies, in particular their incline towards techno optimism or pessimism, as well as their attitudes and engagement with AI, the participants of the survey were presented with a set of five statements to evaluate (see appendix 1). The first statement concerned the impact of technology on society. The results show that nearly 53% of the surveyed individuals express positive or very positive views towards the impact of the technology, while a considerable share of respondents, making up approximately 38%, hold a more neutral stance (see figure 2).

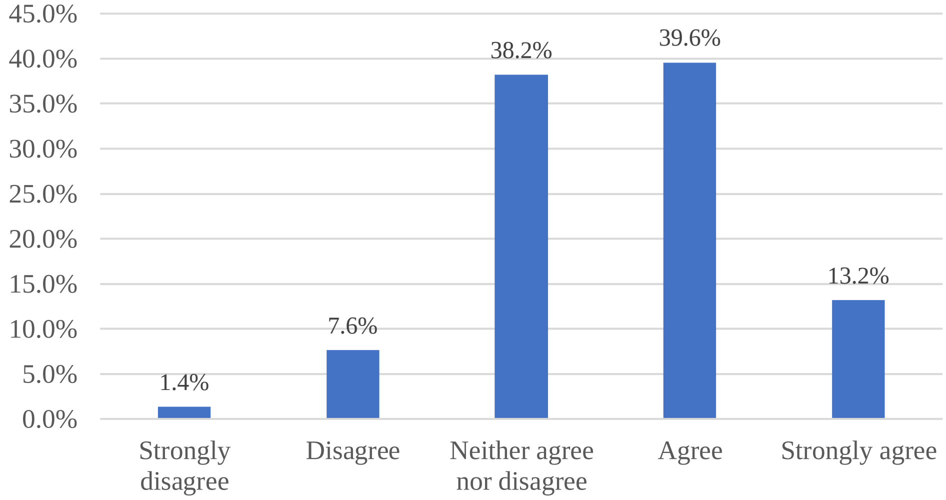


Figure 2. Evaluation of positive impact of technology on society (n=144).
Source: own elaboration

Surprisingly, only 9.3% of respondents perceived the impact of technology as negative. These results indicate a prevailing techno-optimism among the respondents. Furthermore, a Chi-squared test was conducted to examine the relationships between various socio-demographic parameters and these results. No significant differences were observed in terms of gender, age groups, or education levels, suggesting that the perception of the impact of technology as positive or negative was consistent across all of these groups (see appendix 25). However, a significant relationship was discovered when examining the affiliation of individuals with a potentially vulnerable group and their views. Those individuals belonging to ethnic minorities, LGBTQ+ communities, political activists, or those with disabilities more often than other respondents expressed a strong affirmation of the positive impact of technology on society (see appendix 26).

Regarding the frequency of encountering AI, the majority of respondents are increasingly encountering AI in their daily lives (see appendix 24).

Moving on to the evaluation of attitudes towards AI effectiveness in decision-making, in particular considering some superiority over human capabilities, and the perspectives of individual trust in AI-made decisions, the respondents evaluations are graphically presented (see figures 3 and 4).

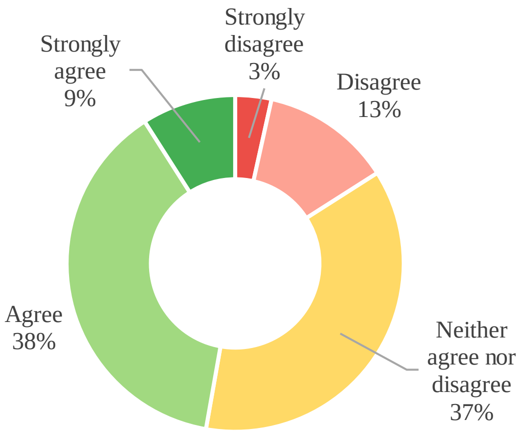


Figure 3. Evaluation of AI making more effective decisions (n=144).
Source: own elaboration

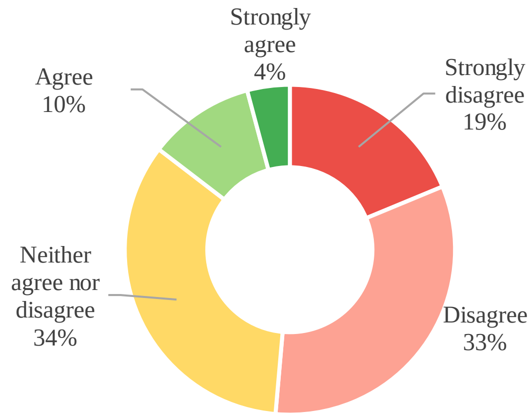


Figure 4. Evaluation of having full trust in AI decisions (n=144).
Source: own elaboration

While only 16% of the respondents believed that AI could not make more effective decisions than humans, the majority, or 48%, exhibited strong confidence in AI's superiority in decision-making capabilities regarding effectiveness (see figure 3). However, a contrasting perspective emerges when looking at trust in AI decisions on an individual level. Only 14.6% of respondents indicated full or close to full trust in AI decisions, with the majority of respondents highlighting a lack of full trust in them (see figure 4). Furthermore, no significant differences were found among various age groups, genders, education levels, or affiliations with potentially vulnerable groups, indicating that the results for trust in AI decisions are consistent across all of these groups (see appendix 27).

Shifting focus to the respondent's perception of AI in terms of bias, the majority of respondents (approximately 35%) agree that AI can have prejudice towards certain groups of people (see figure 5).

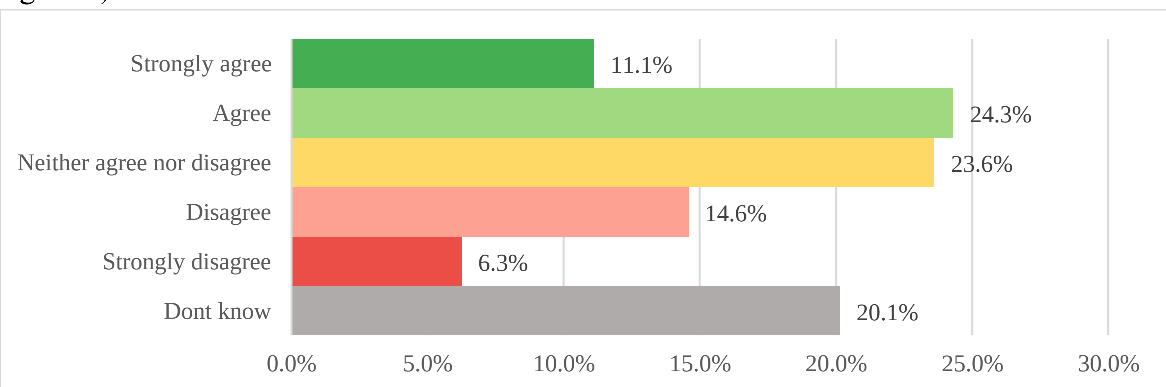


Figure 5. Evaluation of possibilities for AI to have biases (n=144).
Source: own elaboration

The remaining respondents are split between holding a neutral opinion, disagreeing with the idea of AI having bias, and not having a definite opinion on the topic (see figure 5).

Furthermore, when asked to evaluate their overall confidence in these answers, 67% of respondents stated that they were confident or extremely confident about their evaluation. It is noteworthy that respondents who expressed more radical opinions tended to be more confident in their responses (see appendix 29). When looking at the evaluation differences, no significant variances were found between respondents representing various age groups, genders, education levels, or affiliations with potential vulnerable groups (see appendix 28).

Turing to the application for loan data, 43.1% of those surveyed have applied for a loan at some point in their lives, with 43.5% of them experiencing a denial (see appendix 30). A noticeable trend emerges, indicating that the likelihood of individuals having applied for a loan increases the older they are (see appendix 31). Besides, older individuals tend to have fewer instances of loan denials (appendix 32).

When investigating the awareness of respondents about alternative credit scoring, the majority of them, 67%, indicated that they had never heard of such a creditworthiness evaluation method, while 22% pointed out that they were not sure about it (see appendix 33).

Following the evaluation of awareness regarding alternative credit scoring, respondents were presented with a scenario in which they are applying for a loan but eventually get denied due a to lack of credit history (see appendix 1). The bank offers to perform an alternative creditworthiness evaluation, which involves sharing personal non-financial information such as demographic data, including employment and education, telco data, geolocation, a report on selected utility payments over the course of the past 3 years, and access to some of your social media analytics. In response to this scenario, 41% of respondents indicated that they would agree or would likely agree to undergo assessment through the alternative credit evaluation method (see figure 6).

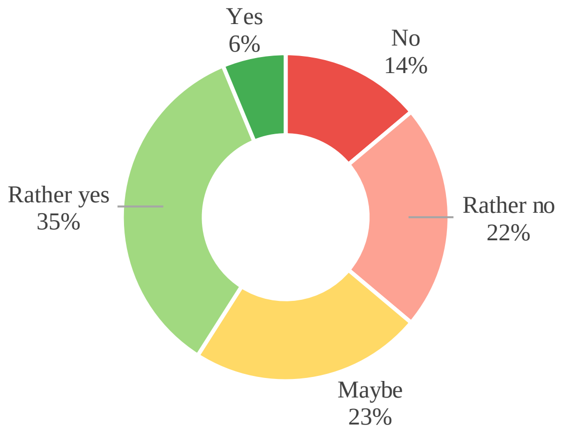


Figure 6. Participation in alternative creditworthiness assessment based on the offered scenario (n=144).
Source: own elaboration

Nevertheless, a comparable proportion of respondents, composing 36%, would definitely not or rather not agree with such a evaluation. Notably, no statistically significant differences were observed across various age groups, gender groups, education levels, or places of residence (see appendix 34). Interestingly, individuals, who express beliefs in the superior decision-making capabilities of artificial intelligence were more inclined to agree to be evaluated under this method (see appendix 35).

For those respondents who selected answers "no" or "rather no," an additional question was presented to explore the reasons behind their hesitance to agree to alternative credit scoring evaluations. Investigation revealed that the most common reasons, in 62% of cases, was the unwillingness to disclose the necessary information (demographic data, mobile activity data, location data, utility payment data, access to social media analytics) (see appendix 36). Another significant reason indicated by 44% was overall distrust in the method of evaluation, and 29% expressed distrust in AI. Surprisingly, only 13% indicated their fear of discrimination (appendix 36).

Next, respondents were asked to evaluate five statements concerning the implications and selected aspects of the alternative credit scoring method on a scale from one to five, where one indicated strong disagreement and five indicated strong agreement. Analysing the evaluations provided by the respondents, it is apparent that when it comes to the belief in enhanced precision, which is a result of non-financial data incorporation, mainly the one based on the behavioural pattern's direct or indirect reflections, the respondents encountered difficulty in unequivocally evaluating these effects, rating them on average with a coefficient of 3.1 (see figure 7).

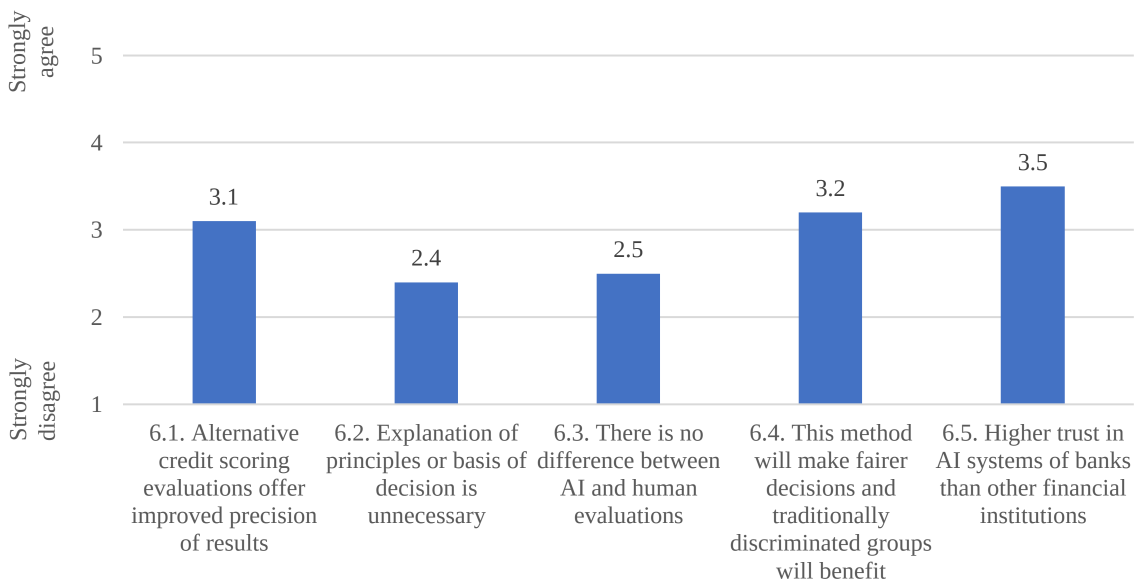


Figure 7. Results from evaluation of various statements regarding alternative creditworthiness method (n=144).
Source: own elaboration

However, nearly one-third of the respondents indicated that the outcomes would be more precise (see appendix 38). Upon examining the results, a positive correlation was discovered between the frequency of encountering AI in every life and trust in the opportunities of the method for more precise evaluation (see appendix 39).

Next, the comfort levels of respondents in situations where the principles of the method or reasons behind a decision were not explained were evaluated, resulting in an average coefficient of 2.4. This indicates that it is important for respondents to receive an explanation and understand the underlying principles of the alternative credit worthiness evaluation calculations (see figure 6). More than half of the respondents, or 57%, indicate that explanation and understanding of its principles are increasingly important for them (see appendices 37 and 38).

When examining the importance of the way credit worthiness is assessed, whether by AI or a human, respondents tend to disagree that there is no difference in the assessment process. Similarly, in line with the evaluation of the first statement regarding improved precision, respondents have difficulty clearly assessing the possibility of fairer decisions and benefits for groups of people who are traditionally discriminated against within society (see figure 7). When looking at the evaluation differences between respondents representing various age groups, genders, education levels, or affiliations with potential vulnerable groups, no significant variances were found (see appendix 40).

Lastly, in examining whether respondent trust in AI systems varies depending on the institutions where they are implemented, the study revealed that respondents generally exhibit higher levels of trust in the AI systems of banks compared to other financial institutions or credit information bureaus (see figure 7).

Furthermore, when assessing the extent to which respondents are willing to receive various types of information, it was determined that their willingness is relatively limited. Among the different types of information provided, the average respondents displayed only a mild willingness to share demographic data such as gender, age, education level, and family status, evaluating it with a coefficient of 3.8 (see figure 8).

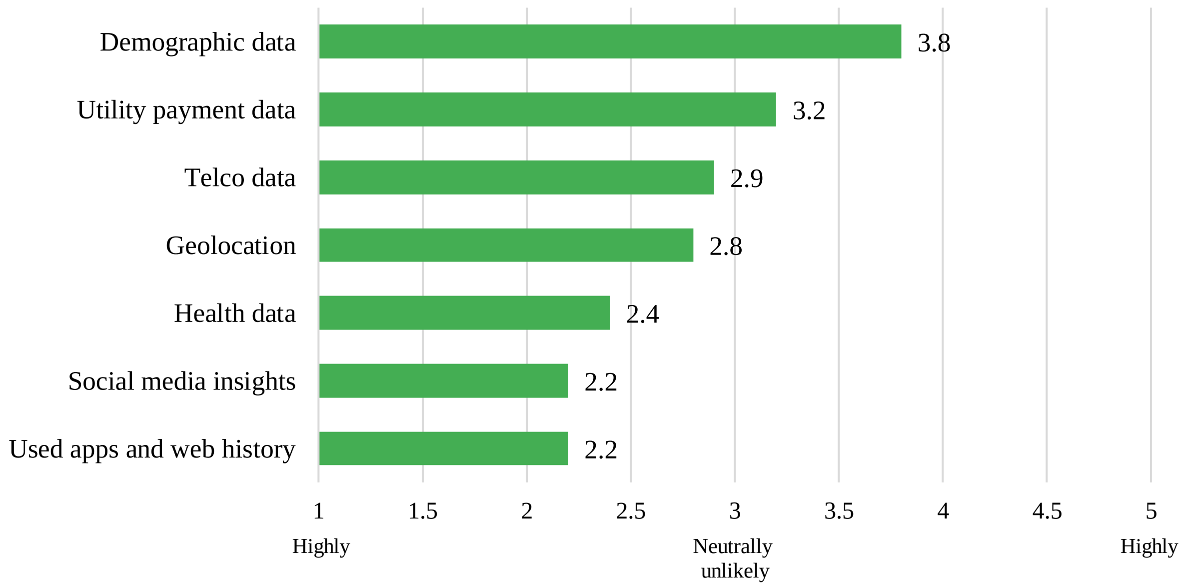


Figure 8. Willingness of respondents to share various types of information for creditworthiness evaluation (n=144).
Source: own elaboration

Additionally, respondents exhibited a certain inclination to share information as utility payments (3.2), data from their mobile carriers (2.9), and geolocation (2.8), however their level of eagerness to do so was rather neutral (see figure 7). On the other hand, respondents expressed even less willingness to provide data on their health status (2.4), social media insights (2.2), and apps used on their mobile devices and activities on the web (2.2) (see figure 7). Among all the types of data evaluated, the assessments provided by respondents showed rather high variations, especially in relation to geolocation and utility payment data (see appendix 42). When examining the evaluation differences between respondents representing potentially vulnerable groups, no significant variances were found (see appendix 41).

Besides, when evaluating respondents' willingness to install an app designed to track their activities, nearly four out of five respondents declared that they would not be ready for such action (see appendix 43).

In terms of assessing the comfort levels of respondents when it comes to sharing sensitive personal information with financial institutions, on average, they expressed indifference towards sharing information on sexual orientation, religious affiliation, gender identity, and membership in associations (see figure 9).

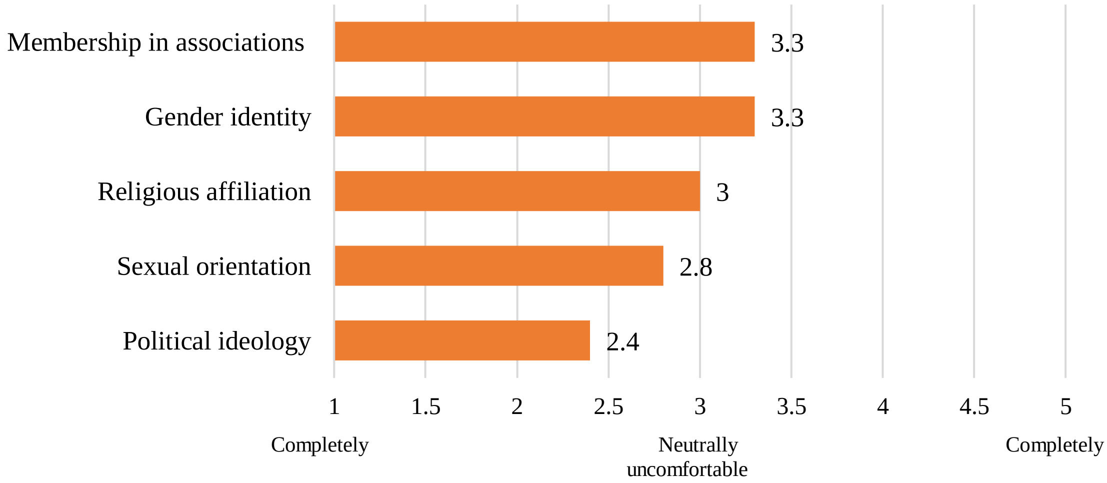


Figure 9. Comfort levels of respondents to share various types of sensitive information with financial institutions (n=144).

Source: own elaboration

Among the parameters evaluated, the only personal information that respondents, on average, would not feel comfortable sharing their political ideology (see figure 9). The results from this evaluation show higher variations compared to the data type evaluation. Sexual orientation, gender identity, and religious affiliation were the sensitive information areas where the largest variations between the respondents occurred (see appendix 44). Despite the average indifference towards sharing information about sexual orientation, the majority of respondents indicated complete discomfort with sharing this type of information (see appendix 44). Similarly, sharing political ideology was recognised as a type of sensitive information that the majority of respondents would feel completely uncomfortable sharing with financial institutions (see appendix 44).

When evaluating the main concerns that the respondents might have when interacting and sharing of information with AI systems, it was found that the majority, approximately three out of five people, or 72.9%, have a fear of information leaks (see figure 10).

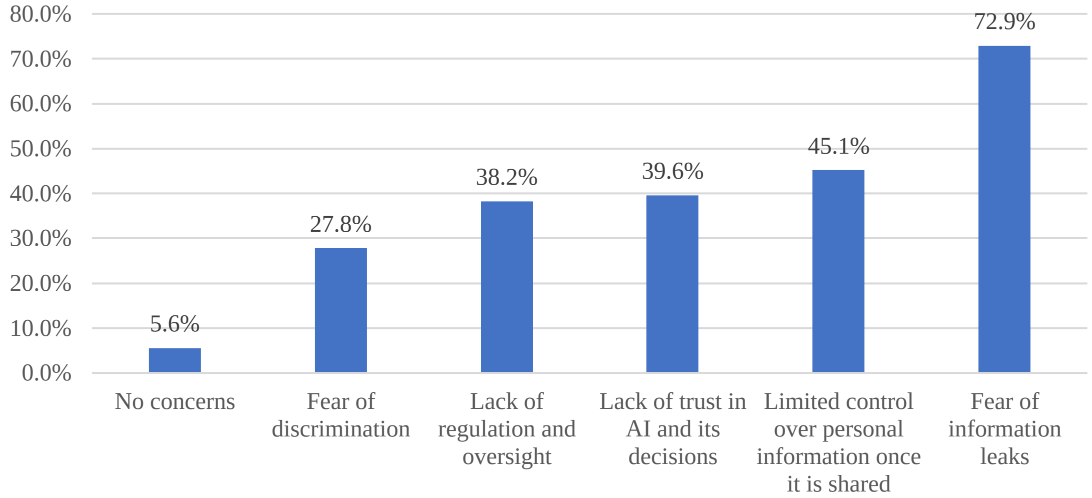


Figure 10. Main concerns of respondents regarding interaction and sharing of information with AI systems (n=144).

Source: own elaboration

Looking at the evaluation differences between respondents representing various age groups, genders, education levels, or affiliations with potential vulnerable groups, it was found that fear of information leaks varies significantly between the genders (see appendix 45). While both men and women expressed concerns about information leaks, women tended to indicate this concern more frequently than men (see appendix 46).

Another notable concern, although not as common as the fear of information leaks, is the lack of control over personal information once it is shared with AI, which affects 45.1% of respondents. Additionally, on average, two in five people present a general lack of trust in AI and its decisions, and in 38.2% of cases, there is a concern over sufficient supervision or normative regulations regarding AI (see figure 10).

Only 27.8% expressed fear of the discrimination that the sensitive factors provide. Surprisingly, no statistically significant differences were identified between respondents representing various age groups, genders, education levels, or affiliations with potential vulnerable groups (see appendix 47). Among all respondents, only 5.7% indicated that they do not have any concerns about sharing information with AI algorithms (see figure 10).

Moving on, in an attempt to determine whether the risk of discrimination outweighs the benefits provided by alternative credit scoring and whether the unwillingness to agree to such evaluation stems from a distrust in AI or concerns about its flaws, respondents were divided into two groups and asked two slightly modified questions (see appendix 1). The results reveal that overall, respondents would rather not agree to be evaluated under such a method if there is a risk of discrimination, despite the potential gains (see figure 11).

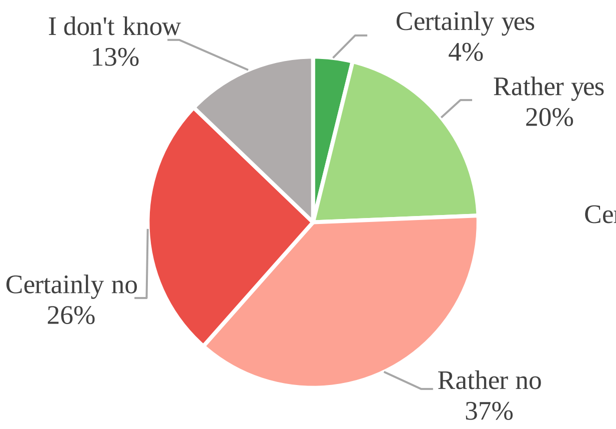


Figure 11. Participation in alternative credit scoring if there is a risk of bias (n=144).
Source: own elaboration

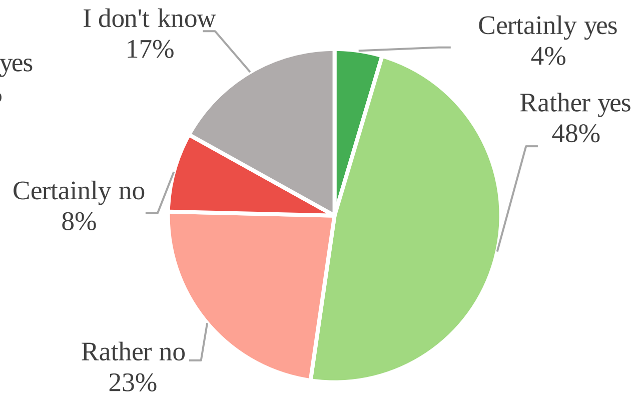


Figure 12. Participation in alternative credit scoring if there is no risk of bias (n=144).
Source: own elaboration

The results indicate that there are no statistically significant differences between respondents representing various age groups, genders, education levels, or affiliations with potential vulnerable groups (see appendix 48).

However, when the risk of discrimination is absent and the significant gains are still preserved, in the majority of cases, respondents would rather agree to be evaluated under such a method (see figure 12). The results show statistically significant differences between respondents representing different age and community groups (see appendix 49). Respondents who identified themselves as belonging to communities in the capital city, capital city metropolitan area, or rural areas and who currently reside outside of the territory of Latvia, were more inclined than average to participate in alternative credit scoring evaluations if they were bias-free (see appendix 50). In terms of age, respondents aged 22–25 and 30–41 showed greater inclination to participate in alternative credit scoring than the rest of the respondents, who predominantly represent individuals younger than 22 and those who were 42 years old and older (see appendix 51).

The aim of the survey was to investigate the perceptions of AI, trust in its decisions, and attitudes towards sharing non-financial personal data among the general population in Latvia. The results show that while the majority of respondents express positive views on the impact of the technology and, overall, can be considered techno-optimists, they exhibit a lack of full trust in automated decisions made by AI. Furthermore, a significant proportion believed that AI decisions can be bias. Regarding the alternative credit scoring evaluations, respondents show limited willingness to share certain types of personal information. Respondents are not willing to share types of information concerning health status, social media insights, and activities on their mobile devices, or in any way disclose information about their sexual orientation or political ideology. Only slightly more than half of respondents are considering participation in alternative credit scoring if there is no risk of bias, thus indicating that their unwillingness for such evaluations is based on other factors than bias. A significant proportion of respondents expressed information leaks as their primary concern when interacting with AI systems.

5. DISCUSSION

The aim of this master thesis was to examine the effects of use of AI for financial services, evaluate the possibility of algorithmic bias presence in AI algorithms for credit scoring and determine its consequences for stakeholders.

5.1. The simulation

Assessing the susceptibility to biases is of rising importance due to the increasing presence of machine learning algorithms in the sector of finance (Xie, 2019; Aziz et al., 2022). Credit scoring practises have undergone a significant transformation with the emergence of machine learning algorithms, allowing automation of the processes, and enabling the integration of non-financial data for creditworthiness evaluations (Njuguna & Sowon, 2021). Therefore, the main objective of this research is to investigate how the use of non-financial data in AI-based credit scoring algorithms data influences their bias.

Previous scientific literature on AI based credit scoring using non-financial data has primarily focused on the aspect of enhancements of the accuracy indicators for various machine learning techniques, studied by Khemakhem & Boujelbene (2018), Hindistan et al. (2019), and Djeundje et al. (2021), besides finding that incorporation of non-financial data increases prediction accuracy for credit scoring systems. However, only a few papers have explored the risks of bias associated with using non-financial data, mostly from a theoretical perspective. Findings by Purda & Ying (2022) and Packin & Lev-Aretz (2018) suggest that credit scoring algorithms using non-financial data might exhibit prejudice against certain societal groups due to bias contained in the used datasets or the design of algorithms reflecting human bias. Furthermore, Kelly & Mirpourian (2021) tested credit scoring algorithms for bias using synthetic data and identified unfair treatment of women applicants as a result of intentionally using flawed data.

Building on the identified connection between non-financial data and unwanted discrimination, my findings, through a simulation exercise of Bhatnagar's (2023) credit approval prediction algorithm trained on real-world non-financial data, closely align with those of Kelly & Mirpourian (2021). I discovered that non-financial data, specifically in this study, socio-demographic and closely related data, can introduce biases by providing unequal treatment to certain groups as women, presumably due to unbalanced training data. My results consistently show gender discrimination, with women receiving lower confidence scores and higher denial rates compared to men. Male profiles are identified as being 2.5 times more likely to receive a positive prediction than female profiles (see appendix 21). Additionally, my study indicates the presence of other unequal treatment factors, as those associated with certain characteristics of family status (e.g., being single or without children), occupation status (low prestige jobs), and education level (secondary education). These factors present lower opportunities for positive prediction by algorithms, potentially limiting equal opportunities for obtaining financial instruments (see Section 4.1). Importantly, these other factors were not studied by the authors mentioned earlier.

However, it is important to acknowledge the limitations of this study. It represents a simplified machine learning credit scoring model that does not include the full scale of parameters such as telco data, geolocation, mobile device activities, or other similar data explored or used in the models of Simumba et al. (2018) or Seon (2023); therefore, the variables reflecting socio-demographic features in this simulation might have a higher impact on the final prediction than in other models. Furthermore, the sample size of profiles tested is limited, consisting of only 126 profile tests, and was primarily focused on the comparison of male and female profiles. Furthermore, data used in the simulation is derived from an unspecified geographical location, which poses challenges in terms of interpreting the results and understanding the context and background of the data.

Overall, the study highlights the potential risks of incorporating non-financial data into credit scoring algorithms, as it can perpetuate biases and hinder equal access to financial instruments. Therefore, financial institutions implementing non-financial data should carefully consider the benefits and risks associated with incorporating such features into their credit scoring algorithms. It is crucial to ensure complete, properly categorised, and balanced training data sets to provide equal representation of the groups within a society and decrease the probability of sampling or labelling bias, which might reflect systematic or inherited bias from previous evaluation systems.

Furthermore, institutions using credit scoring algorithms should implement explainability principles in their algorithms to better understand the factors influencing credit decisions. Regular algorithm testing is also necessary to ensure that unintentional biases do not develop over time.

In this context, it is essential to consider the European Commission's "Artificial Intelligence Act," which, although not yet passed, will directly affect AI-based credit scoring algorithm use in the European Union (EU), primarily by requiring third-party ex-ante and ex-post conformity assessment with internal checks and risk mitigations and by banning trustworthiness evaluations based on an individual's social behaviour (European Commission, 2021). However, such requirements are absent in other regions, in particular Southeast Asia, where the majority of the countries present weak or still emerging regulatory frameworks for data protection and AI utilisation in both public and private environments (Chitturu, Lin, Sneader, Tonby, & Woetzel, 2017; Noor & Manantan, 2022).

Considering these factors, I recommend that financial institutions proactively eliminate the use of data that reflects social behaviour in their credit scoring before the Artificial Intelligence Act comes into effect. It is also important to reconsider the use of non-financial data associated with socio-demographics. Furthermore, I suggest that financial institutions and regulatory bodies develop unified guidelines for the utilisation of non-financial data in credit scoring systems, following the upcoming regulation from the European Commission. Additionally, a unified framework could be established for assessing the fairness and non-discrimination of algorithms used in the financial sector. Lastly, regulatory institutions should expand their oversight responsibilities for institutions utilising AI credit scoring systems to ensure fair and non-discriminatory credit scoring practises.

5.2. The survey

The final objective of this master's thesis is to explore the perceptions of AI, trust in its decisions, and attitudes towards sharing non-financial personal data among the general population of an EU country. This exploration is crucial to understand how the population perceive AI algorithms and to assess the factors that trigger trust and acceptance of their use.

The survey conducted in this study reveals that despite positive attitudes towards technological development and the association of the population with techno-optimism, the population exhibits a lack of trust in decisions made by AI. This low level of trust aligns with previous observations made by Vasiljeva et al. (2021) and the Baltic International Bank (2018). A possible explanation for such a lack of trust in AI decisions could be the perception of a lack of accountability. While humans can be held responsible for their actions, AI algorithms may seem faceless and challenging to assign responsibility to, which contributes to the distrust in AI systems. Negative experiences with AI systems in daily life may also contribute to this scepticism. Furthermore, the results from the survey indicate that the low trust in AI decisions is also linked to a reluctance to be evaluated under alternative credit scoring and scepticism regarding the benefits of alternative credit scoring for fairer decisions, in particular for traditionally discriminated groups.

Regarding the willingness to share different types of information, respondents generally exhibit a neutral stance, indicating only a willingness to share the information if it is necessary. The population analysed shows readiness to share information regarding demographics, utility payments, telco, and geolocations. However, they expressed discomfort when it came to sharing health status, social media insight, and app history, or mobile device activity data, and expressed a dismissive attitude towards installing any tracking software. It was surprising to find that the population was more comfortable sharing telco data, which includes information on the behaviour of individuals when using mobile services, considering it is somewhat similar to sharing data on mobile device activities. Interestingly, no significant differences in sharing various types of data were found between the respondents representing potentially vulnerable groups and other respondents.

Another intriguing finding is that when evaluating the comfort levels of sharing sensitive personal information as sexual orientation, religious affiliation, or gender identity, the population did not express reluctance to do so. This finding is noteworthy considering that the country's population holds socially conservative views (Dimdins, Sandgren, & Montgomery, 2016). Out of all the sensitive personal information types analysed, respondents found sharing their political ideology to be the only inappropriate type of information to share with financial institutions. This suggests that respondents do not perceive factors as sexual orientation, religious affiliation, or gender identity as contributing to discrimination in accessing financial instruments. It may indicate that respondents believe there is no bias against these members of different groups in the given population. However, the results indicate that political views might impact their chances of accessing financial instruments.

The main concern expressed by all respondents regarding interaction with AI systems was the fear of leaks of information individuals share. This indicates a significant level of concern within the population regarding the safety of their information, highlighting the need to address data privacy concerns to promote trust in AI systems, particularly in the context of alternative credit scoring.

It is important to acknowledge the limitations of this attitudes study, which primarily include a relatively small sample size, an overrepresentation of female respondents, and the limitations of the study to one EU country. Therefore, this data might not be representative of the entire EU population and might not be generalised to other regions. Thus, it would be valuable to expand the study to other geographic areas, particularly Southeast Asian countries that show significant potential for financial inclusion through alternative credit scoring algorithms. Additionally, future studies could consider examining attitudes based on different income levels or the degree of necessity for financial instruments to understand how these factors affect perceptions of alternative credit scoring and readiness to share the necessary data.

From a practical implication standpoint, in addition to the implications concerning the simulation part of the study, the survey study underscores the importance of addressing data privacy. Private organisations utilising AI algorithms and especially governmental agencies should prioritise implementing robust data protection measures. Furthermore, financial companies utilising alternative credit scoring should carefully consider the necessary information for their models and find the best balance between the necessary information and respecting the privacy preferences of their clients (existing or potential). The willingness of the population to share certain types of information, such as demographic data and utility payments, can guide financial institutions in focusing on the collection and analysis of this specific data.

When considering the challenges associated with implementing alternative credit scoring systems, several factors may contribute to their emergence. Firstly, there is limited awareness and understanding among the general population regarding AI-based credit scoring principles, specifically automated decision-making. This lack of understanding fosters distrust and a reluctance to participate in such evaluations. Institutions can tackle this issue by organising educational programmes and campaigns that inform individuals about the implications of AI systems and various credit scoring practises, ensuring transparency, and building trust. Secondly, individuals often express a sense of lack of control over their personal information once it is shared with AI systems. Institutions should address these concerns by implementing mechanisms that enable individuals to maintain control over the data they share. Additionally, they should provide a clear and understandable process for requesting the erasure of personal data, as determined by Article 17 of the GDPR (2016), guaranteeing the right to be forgotten.

Considering the findings, it is recommended that financial institutions provide clear and understandable explanations of how AI algorithms make decisions, the types of data used, and how discriminations in algorithms are addressed. Furthermore, researchers and experts in the field should investigate the factors, both individual interference and publicly discussed, as well as beliefs or other aspects that influence negative perceptions of AI algorithms.

6. CONCLUSIONS

In light of the increasing presence of machine learning algorithms in the financial sector, this master's thesis aimed to examine the effects of the use of customer-oriented AI systems for financial services by evaluating the possibility of algorithmic bias in alternative credit scoring algorithms and determining its consequences for stakeholders.

The findings reveal that the use of non-financial data in credit scoring algorithms contributes to the risk of biased results. Although it is not a precondition for discriminatory results, as the presence of bias in machine learning models is highly dependent on the characteristics of the training data. Results from the simulation show unequal treatment of female profiles, making them 2.5 times less likely to get approval for a financial instrument than male applicants. Insights into the perspectives of stakeholders concerning alternative credit scoring algorithms show that individuals who exhibit trust in AI decision-making show interest and willingness to be evaluated under such an approach. Overall, respondents show readiness to share various types of information, except data on activities from their smart devices, browsing history, or personal health information. Concerns regarding data privacy are one of the main drivers that significantly influence perceptions of alternative credit scoring algorithms when it involves sharing information with artificial intelligence systems. Addressing the issues of unbalanced training data used in prediction algorithms and confronting the concerns over data privacy has the potential to foster greater approval of AI-powered systems in the financial sector, bringing benefits of competitiveness for financial firms and the potential for inclusion for customers with limited credit histories.

Future research in this field of algorithmic bias in credit scoring algorithms should encompass a wider range of factors to explore how the introduction of additional variables influences the accuracy and fairness of such models. Although more research and simulations are necessary, it is reasonable to assume that the inclusion of non-financial data increases the risk of biased results, although the exact extent of this phenomenon remains unknown, partially due to the lack of a single methodology and virtually limitless variations of alternative credit scoring algorithms (Purda & Ying, 2022), which limits determining any approximations.

From the perspective of attitudes towards AI and alternative credit scoring, there is significant value in broadening the scope of this study to other geographic areas, particularly Southeast Asian countries that exhibit promising potential for financial inclusion through alternative credit scoring algorithms. Besides, it is crucial for researchers and experts in the field to delve into the factors that influence negative perceptions of AI algorithms in decision-making processes, considering both individual perspectives, or beliefs, and publicly raised concerns.

As a whole, the practise of machine learning applications for financial problems goes beyond discussion of them in financial research (Aziz et al., 2022). The results obtained from simulation and the survey contribute to the discussion of ML practises used in alternative credit scoring.

REFERENCE LIST AND BIBLIOGRAPHY

Abdou, H., A. & Pointon, J., (2011), Credit scoring, statistical techniques and evaluation criteria: a review of the literature. In Intelligent Systems in Accounting, Finance and Management, Volume 18, Issue 2-3 p. 59-88. Retrieved from: https://www.doi.org/10.1002/isaf.325 (accessed 14.04.2023);

Aggarwal, N. (2018). Machine Learning, Big Data and the Regulation of Consumer Credit Markets: The Case of Algorithmic Credit Scoring. Discrimination. Retrieved from: http://dx.doi.org/10.2139/ssrn.3309244 (accessed 28.05.2023);

Aji, N. A., & Dhini, A. (2019, July). Credit scoring through data mining approach: A case study of mortgage loan in Indonesia. In 2019 16th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1-5). IEEE. Retrieved from: https://doi.org/10.1109/ICSSSM.2019.8887731 (accessed 10.06.2023);

Asian Development Bank, (2023), “How Alternative Credit Scoring Can Improve the Poor's Access to Loans”, retrieved from: https://development.asia/explainer/heres-how-alternative-credit-scoring-can-improve-poors-access-loans (accessed 09.03.2023);

Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2022). Machine learning in finance: A topic modeling approach. European Financial Management, 28(3), 744-770. Retrieved from: https://doi-org.sire.ub.edu/10.1111/eufm.12326 (accessed 12.05.2023);

Baltic International bank, (2018), “Baltic International Bank Latvijas barometrs”. Retrieved from: https://www.bib.eu/uploads/2017/02/Baltic-International-Bank-Latvijas-barometrs-tehnologijas-02.2018.pdf (accessed 20.05.2023);

Bhatnagar, G., (2023), “CreditCardApprovalPrediction_Pycarett”, retrieved from: https://www.kaggle.com/code/gauri1996/creditcardapprovalprediction-pycarett/notebook (accessed 10.04.2023);

Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25, 197-227. Retrieved from: https://doi.org/10.1007/s11749-016-0481-7 (accessed 10.04.2023);

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32. Retrieved from: https://doi.org/10.1023/A:1010933404324 (accessed 02.05.2023);

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. Computational Economics, 57, 203-216. Retrieved from: https://doi.org/10.1007/s10614-020-10042-0 (accessed 14.05.2023);

Central Statistical Bureau of Latvia (2023), “Population and population change”. Retrieved from: https://stat.gov.lv/en/statistics-themes/population/population/247-population-and-population-change (accessed 30.05.2023);

Chitturu, S., Lin, D., Y., Sneader, K., Tonby, O., & Woetzel, J., (2017), “Artificial intelligence and Southeast Asia’s future”. Retrieved from: https://www.mckinsey.com/nl/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/AI%20and%20SE%20ASIA%20future/Artificial-intelligence-and-Southeast-Asias-future.pdf (accessed 16.05.2023);

Dimdins, G., Sandgren, M. & Montgomery, H. (2016). Psychological variables underlying political orientations in an old and a new democracy: A comparative study between

Sweden and Latvia. Scandinavian Journal of Psychology, 57, 437– 445. Retrieved from: https://doi.org/10.1111/sjop.12314 (accessed 13.06.2023);

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in Finance (Vol. 1170). Berlin and Heidelberg: Springer International Publishing, retrieved from: https://doi.org/10.1007/978- 3- 030- 41068- 1_1 (accessed 10.05.2023);

Djeundje, V., B., Crook, J., Calabrese, R. & Hamid, M., (2021), Enhancing credit scoring with alternative data. In Expert Systems with Applications, 163, 113766, retrieved from: https://doi.org/10.1016/j.eswa.2020.113766 (accessed 12.04.2023);

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? (pp. 3-11). Springer International Publishing. Retrieved from: https://doi.org/10.1007/978-3-319-18305-3_1 (accessed 11.05.2023);

European Commission (2021), “Proposal for a regulation of the European parliament and of the council: Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts”. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (accessed 17.05.2023);

FRA, (2022), “Bias in Algorithms – Artificial Intelligence and Discrimination”, retrieved from: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf (accessed 24.02.2023);

GDPR (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 199, 1-8. Retrieved from: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed 14.05.2023);

Gogas, P., & Papadimitriou, T. (2021). Machine learning in economics and finance. Computational Economics, 57, 1-4. Retrieved from: https://doi.org/10.1007/s10614-021-10094-w (accessed 10.05.2023);

Google, (2022a), “Classification: Accuracy”, retrieved from: https://developers.google.com/machine-learning/crash-course/classification/accuracy (accessed 05.06.2023);

Google, (2022b), “Classification: ROC Curve and AUC”, retrieved from: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc (accessed 05.06.2023);

Google, (2022c), “Precision and Recall”, retrieved from: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall (accessed 05.06.2023);

Gsenger, R. & Strle, T., (2021), Trust, automation bias and aversion: algorithmic decision-making in the context of credit scoring. In Interdisciplinary Description of Complex Systems: INDECS, 19(4), 540-558. Retrieved from: https://www.doi.org/10.7906/indecs.19.4.7 (accessed 19.04.2023);

GSMA (2021), “The Mobile Economy Asia Pacific 2021”. Retrieved from: https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/08/GSMA_ME_APAC_2021_Web_Singles.pdf (accessed 13.05.2023);

GSMA, (2022), "The Mobile Economy 2022". Retrieved from: https://www.gsma.com/mobileeconomy/wp-content/uploads/2022/02/280222-The-Mobile-Economy-2022.pdf (accessed 18.05.2023);

Hindistan, Y. S., Kıyakoğlu, B. Y., Rezaeinazhad, A. M., Korkmaz, H. E., & Dağ, H., (2019), Alternative Credit Scoring and Classification Employing Machine Learning Techniques on a Big Data Platform. In 4th International Conference on Computer Science and Engineering (pp. 1-4), retrieved from: https://www.doi.org/10.1109/UBMK.2019.8907113 (accessed 18.04.2023);

Huang, Y. P., & Yen, M. F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. Applied Soft Computing, 83. Retrieved from: https://doi.org/10.1016/j.asoc.2019.105663 (accessed 11.05.2023);

Kelly, S., & Mirpourian, M., (2021), "Algorithmic Bias, Financial Inclusion, and Gender", retrieved from: https://www.womensworldbanking.org/wp-content/uploads/2021/02/2021_Algorithmic_Bias_Report.pdf (accessed 23.02.2023);

Khemakhem, S., & Boujelbene, Y. (2018). Predicting credit risk on the basis of financial and non-financial variables and data mining. Review of Accounting and Finance. Retrieved from: https://doi.org/10.1108/RAF-07-2017-0143 (accessed 15.06.2023);

Lim, K., J., (2022), "How to close Southeast Asia's financial inclusion gap", retrieved from: https://www.weforum.org/agenda/2022/02/closing-southeast-asia-s-financial-inclusion-gap/ (accessed 13.05.2023);

Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. Expert Systems with Applications, 42(10), 4621-4631. Retrieved from: https://doi.org/10.1016/j.eswa.2015.02.001 (accessed 8.05.2023);

Maple Tech International, (2022), "Sample size calculator", retrieved from: https://www.calculator.net/sample-size-calculator.html?type=2&cl2=95&ss2=144&pc2=50&ps2=1883008+&x=84&y=20#findci (accessed 30.05.2023);

Mhlanga, D. (2020). Industry 4.0 in finance: the impact of artificial intelligence (ai) on digital financial inclusion. International Journal of Financial Studies, 8(3), 45. Retrieved from: https://doi.org/10.3390/ijfs8030045 (accessed 12.05.2023);

Neudert, L. M., Knuutila, A., & Howard, P. N. (2020). Global attitudes towards AI, machine learning & automated decision making. Working paper 2020.10, Oxford Commission on AI & Good Governance. Retrieved from: https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/11/2020/10/GlobalAttitudesTowardsAIMachineLearning2020.pdf (accessed 20.05.2023);

Njuguna, R., & Sowon, K. (2021, June). Poster: A scoping review of alternative credit scoring literature. In ACM SIGCAS Conference on Computing and Sustainable Societies (pp. 437-444). Retrieved from: https://doi.org/10.1145/3460112.3471972 (accessed 15.06.2023);

Noor, E., & Manantan, M., B., (2022), “Raising Standards: Data and artificial intelligence in Southeast Asia”. Retrieved from: https://asiasociety.org/sites/default/files/inline-files/ASPI_RaisingStandards_report_fin_web_0.pdf (accessed 16.05.2023);

OECD (2021), “Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers”. Retrieved from: https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm (accessed 15.05.2023);

Packin, N. G., & Lev-Aretz, Y. (2018). Learning algorithms and discrimination. In Research handbook on the law of artificial intelligence (pp. 88-113), retrieved from: https://doi.org/10.4337/9781786439055.00014 (accessed 08.03.2023);

Pethig, F. & Kroenung, J., (2022), Biased Humans, (Un)Biased Algorithms. In Journal of Business Ethics volume 183, pages 637–652. Retrieved from: https://doi.org/10.1007/s10551-022-05071-8 (accessed 21.04.2023);

Popovych, B., (2022), Application of AI in Credit Scoring Modeling. In Springer Fachmedien Wiesbaden GmbH, retrieved from: https://doi.org/10.1007/978-3-658-40180-1 (accessed 13.04.2023);

Purda, L., & Ying, C. (2022). Consumer Credit Assessments in the Age of Big Data. In Big Data in Finance: Opportunities and Challenges of Financial Digitalization (pp. 95-113), retrieved from: https://doi.org/10.1007/978-3-031-12240-8 (accessed 24.02.2023);

PWC (2023), Understanding algorithmic bias and how to build trust in AI, retrieved from: https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html (accessed 18.04.2023.);

Seon, (2023), Alternative Credit Scoring: What is it and How it Works, retrieved from: https://seon.io/resources/guides/alternative-credit-scoring/#h-what-is-alternative-credit-scoring (accessed 14.04.2023);

Simumba, N., Okami, S., Kodaka, A. & Kohtake, N., (2018), Alternative Scoring Factors using Non-Financial Data for Credit Decisions in Agricultural Microfinance. In IEEE International Systems Engineering Symposium (ISSE) (pp. 1-8), retrieved from: https://www.doi.org/10.1109/SysEng.2018.8544442 (accessed 12.04.2023);

Song, X., (2020), “Credit Card Approval Prediction”, retrieved from: https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction (accessed 10.04.2023);

Statista, (2023a), “Penetration rate of smartphones in Europe from 2013 to 2028”. Retrieved from: https://www-statista-com.sire.ub.edu/forecasts/1147144/smartphone-penetration-forecast-in-europe (accessed 18.05.2023);

Statista, (2023b), “Penetration rate of smartphones in Romania from 2013 to 2028”. Retrieved from: https://www-statista-com.sire.ub.edu/forecasts/1145357/smartphone-penetration-forecast-in-romania (accessed 18.05.2023);

Statista, (2023c), “Penetration rate of smartphones in Bulgaria from 2013 to 2028”. Retrieved from: https://www-statista-com.sire.ub.edu/forecasts/1146748/smartphone-penetration-forecast-in-bulgaria (accessed 18.05.2023);

The World Bank (2021a), "The Global Findex Database 2021". Retrieved from: https://www.worldbank.org/en/publication/globalfindex/interactive-executive-summary-visualization (accessed 12.05.2023);

The World Bank (2021b), Global Financial Inclusion, retrieved from: https://databank.worldbank.org/source/global-financial-inclusion (accessed 14.04.2023);

The World Bank, (2023), "Financial Inclusion", retrieved from: https://www.worldbank.org/en/topic/financialinclusion/overview (accessed 12.05.2023);

Uddin, M. S., Chi, G., Al Janabi, M. A., & Habib, T. (2022). Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability. International Journal of Finance & Economics, 27(3), 3713-3729. Retrieved from: https://doi-org.sire.ub.edu/10.1002/ijfe.2346 (accessed 02.05.2023);

United Nations (2023), "Reduce inequality within and among countries", retrieved from: https://sdgs.un.org/goals/goal10 (accessed 08.03.2023);

Vasiljeva, T., Kreituss, I., & Lulle, I. (2021). Artificial intelligence: the attitude of the public and representatives of various industries. Journal of Risk and Financial Management, 14(8), 339. Retrieved from: https://doi.org/10.3390/jrfm14080339 (accessed 20.05.2023);

Vered, M., Livni, T., Howe, P. D. L., Miller, T., & Sonenberg, L. (2023). The effects of explanations on automation bias. Artificial Intelligence, 103952. Retrieved from: https://doi.org/10.1016/j.artint.2023.103952 (accessed 14.05.2023);

Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. Management Review Quarterly, 1-41. Retrieved from: https://doi.org/10.1007/s11301-023-00320-0 (accessed 14.05.2023);

Woodward, M., (2023), "Artificial Intelligence statistics for 2023", retrieved from: https://www.searchlogistics.com/learn/statistics/artificial-intelligence-statistics/#:~:text=AI%20technology%20is%20one%20of,a%2026.1%25%20global%20market%20share (accessed 10.05.2023);

Xie, M., (2019), Development of Artificial Intelligence and Effects on Financial System, In Journal of Physics: Conference Series, Volume 1187, Issue 3, retrieved from: https://doi.org/10.1088/1742-6596/1187/3/032084 (accessed 11.04.2023);

Yhip, T. M., Alagheband, B. M., (2020). Credit Analysis and Credit Management. The Practice of Lending: A Guide to Credit Analysis and Credit Risk, 3-46. Retrieved from: https://doi.org/10.1007/978-3-030-32197-0_1 (accessed 20.04.2023);

Zhang, S., & Kuhn, P. (2022). Understanding Algorithmic Bias in Job Recommender Systems: An Audit Study Approach. Retrieved from: https://www.aeaweb.org/conference/2023/program/paper/aak44e3s (accessed 02.04.2023).

LIST OF APPENDICES

Appendix 1. Questionnaire (English version)

Use of AI in credit scoring algorithms

My name is Juris Antonevics (juris.antonevics@ba.lv), and I am a student at the University of Barcelona. As a part of my master's thesis, tutored by asst. prof. Guillem Riambau (griambau@ub.edu), I am conducting a research on the effects of artificial intelligence in financial markets, in particular exploring algorithmic bias and its impact on stakeholders.

I kindly ask you to participate in a brief survey that should take no more than 5 - 7 minutes to complete.

The survey responses will remain anonymous, and will only be used in aggregated form for research purposes.

I appreciate your valuable time and participation in this survey!

Iniciar sesión en Google para guardar lo que llevas hecho. Más información

* Indica que la pregunta es obligatoria

Your age *

- 18-21
- 22-25
- 26-29
- 30-33
- 34-37
- 38-41
- 42 or older

Your gender *

- Male
- Female
- Other

Your education level *

- ◯ Below secondary
- ◯ Secondary
- ◯ Bachelor
- ◯ Master
- ◯ Doctor
- ◯ Otro: ____

Type of community you live in *

- ◯ The capital city
- ◯ Suburban area of the capital city
- ◯ Another large city
- ◯ Small city or town
- ◯ Rural area
- ◯ Otro: ____

Please indicate if you belong to any of the listed categories *

- ☐ Ethnic minority
- ☐ Refugee
- ☐ Person with disabilities
- ☐ LGBTQ+
- ☐ Political activist
- ☐ Prefer not to say
- ☐ None of the above

1. Please evaluate the following statements on general attitudes towards technologies and AI on a scale from 1 to 5.

(1 – Strongly disagree, 2 – Disagree, 3 –Neither agree nor disagree, 4 –Agree, 5 – Strongly agree)

1.1. Impact of technology on society is positive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

1.2. I am increasingly encountering artificial intelligence (AI) in my daily life. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

1.3. AI algorithms can make decisions more efficiently (quicker, with less errors, more streamlined) than humans. *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

1.4. I fully trust decisions made by AI systems. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

1.5. AI algorithms can have bias towards certain groups of people. *

E.g. unfair treatment of one of the genders

- ○ 1. Strongly disagree
- ○ 2. Disagree
- ○ 3. Neither agree, nor disagree
- ○ 4. Agree
- ○ 5. Strongly agree
- ○ I don't know
- ○ Otro: ____

1.6. How confident are you on the previous answer? *

	1	2	3	4	5	
Not confident at all	○	○	○	○	○	Very confident

2. Have you ever applied for a credit? *

- ○ Yes
- ○ No
- ○ Don't want to answer

3. Have you ever been denied credit application? *

If the answer to the previous question was "No", please select "Not applicable".

- ○ Yes
- ○ No
- ○ Don't want to answer
- ○ Not applicable

Alternative credit scoring algorithm definition: A credit worthiness evaluation method which relies on behavioural patterns and non-financial data sources (e.g., demographic, education, telco, social media, and other data) to run AI based prediction algorithms that determine lending trust.

4. Have you heard about alternative credit scoring algorithms? *

- ○ Yes
- ○ No
- ○ I am not sure
- ○ Otro: ____________

Imagine a scenario:

You apply for a loan, but the bank informs you that they have insufficient financial history to perform a traditional creditworthiness evaluation on you. Same situation happens with you in other banks. Under these circumstances one of the banks offers an alternative creditworthiness evaluation that requires you to share personal non-financial information.

The information required:
- demographic data, including your employment, education, marital status,
- mobile data,
- location data,
- report on selected utility payments over the course of the past 3 years,
- access to some data related to your social media.

5.1. In this scenario would you incline to be evaluated under alternative creditworthiness evaluation? *

- ○ Yes
- ○ Rather yes
- ○ Maybe
- ○ Rather no
- ○ No
- ○ Otro: ____________

Siguiente | Borrar formulario

Negative response (additional question)

5.2. Please select the main reason why you do not want to be assessed by this method: *

- ☐ I do not want to share the necessary information
- ☐ Distrust of the method/misunderstanding of its working principles
- ☐ Fear of discrimination
- ☐ Distrust of artificial intelligence and its decisions
- ☐ N/A
- ☐ Otro: ____________

Atrás | Siguiente | Borrar formulario

Second part of the survey

6. Based on the previously addressed scenario please evaluate the following statements on a scale from 1-5:
(1 – Strongly disagree, 2 – Disagree, 3 – Neither agree nor disagree, 4 – Agree, 5 – Strongly agree)

6.1. I believe that evaluating creditworthiness using personal, non-financial based information that directly or indirectly reflects a person's behaviour, will result in a more accurate result. *
(In cases where such information is used in combination with, or independently of, financial information)

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6.2. I am comfortable not being informed about how the AI algorithm determines my creditworthiness and the final decision on my application. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6.3. It doesn't matter if I'm judged by a human or an artificial intelligence. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6.4. I believe that this AI algorithm will make fair decisions that can benefit groups of people who within a society are traditionally discriminated against. *

- ○ Strongly disagree
- ○ Rather disagree
- ○ Neither agree, nor disagree
- ○ Rather agree
- ○ Strongly Agree
- ○ I don't know
- ○ Otro: ________

6.5. I would trust more an AI system of a bank rather than nonbank financial institutions (e.g. microloan organization) or credit information bureau systems. *

- Strongly disagree
- Rather disagree
- Neither agree, nor disagree
- Rather agree
- Strongly Agree
- I don't know
- Otro:

7. Please evaluate how likely you would willingly share the following information for evaluation of your creditworthiness on a scale from 1 to 5. *

	N/A	Highly unlikely	Rather unlikely	Neutrally	Rather likely	Highly likely
7.1. Demographic data (gender, age, race, education, marital status, etc.)	○	○	○	○	○	○
7.2. Mobile data (usage patterns)	○	○	○	○	○	○
7.3. App information and web activity on your smart device	○	○	○	○	○	○
7.4. Activities in social networks (detailed profile analysis)	○	○	○	○	○	○
7.5. Geolocation data (place of birth, current and previous places of residence)	○	○	○	○	○	○
7.6. Information about utility payments	○	○	○	○	○	○

7.7. Health status (medical history)	○	○	○	○	○	○

7.8. Would you be willing to install an app on your phone that tracks your activities for the purpose of alternative credit scoring? *

Activities might include: device use patterns as on-screen time, mobile app usage, web history tracking, access to your contact list and others.

- ○ Yes
- ○ Rather yes
- ○ Rather no
- ○ No
- ○ Don't want to answer
- ○ Otro: ____

Atrás | Siguiente | Borrar formulario

Third part of the survey

In some instances financial institutions that implemented alternative credit scoring have analysed or used the following information: sexual orientation, gender identity, religious affiliation, political ideology, association you belong to.

8. On a scale from 1 to 5 please evaluate how comfortable would you feel sharing above mentioned specific personal information with financial institutions.

	N/A	Completely uncomfortable	Rather uncomfortable	Neutrally	Rather comfortable	Completely comfortable
8.1. Sexual orientation	○	○	○	○	○	○
8.2. Gender identity	○	○	○	○	○	○
8.3. Religious affiliation	○	○	○	○	○	○
8.4. Political ideology	○	○	○	○	○	○

8.5. Associations you belong to (e.g. sports, cultural, professional, and others) ○ ○ ○ ○ ○ ○

9. What would be your main concerns when sharing information with an AI system of financial institution? *

Please select up to 3 concerns

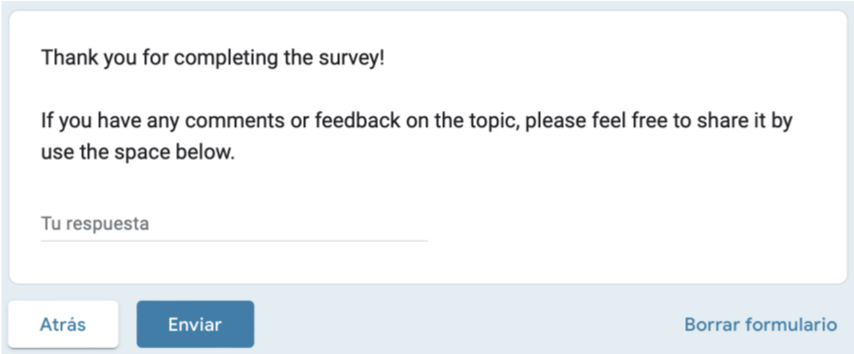
- ☐ Discrimination based on sensitive factors
- ☐ Fear of information leaks
- ☐ Lack of regulation and oversight
- ☐ Lack of trust in AI and its decisions
- ☐ Limited control over personal information once it is shared
- ☐ I don't have any concerns
- ☐ Otro: ______

10.a If it was proven that AI creditworthiness evaluation algorithms can have bias, but on the other hand they reduce application for loan period to just a few days, would you still like to be evaluated under such algorithm? *

- ○ Definitely yes
- ○ Rather yes
- ○ Rather no
- ○ Definitely no
- ○ I don't know
- ○ Otro: ______

10.b If it was proven that AI creditworthiness evaluation algorithms are bias free, and they reduce application for loan period to just a few days, would you like to be evaluated under such algorithm? *

- ○ Definitely yes
- ○ Rather yes
- ○ Rather no
- ○ Definitely no
- ○ I don't know
- ○ Otro: ______



Appendix 2. Description of the testing groups and their profiles for the final testing phase in the simulation task

Testing group	Investigation factors	Characteristics of the profiles				
		Age & Gender	**Marital status & number of children**	**Job position, employment duration & education level**	**Housing type**	**Other features**
1	Influence of the age difference between men and women.	55-, 40-, and 25-year-old men and women	Married with one child	Medium prestige job with various employment durations, having a lower secondary education	Living in a house or apartment	Owning a car, but not extra property. Having a phone and an email.
2	Influence of family status & number of children in comparison between men and women	45-year-old men and women	Various: civil marriage, marriage, separated, single, widow, without children	A medium prestige job with a long employment duration and having a higher education	Living in a house or apartment	Owning a car, but not extra property. Having a phone and an email.
3	Comparison of high, medium, and low prestige job positions & various employment durations between men and women	45-year-old men and women	Married with two children	High prestige job, medium prestige job, low prestige job with a long, medium, or short employment duration and various education levels	Living in a house or apartment	Owning a car, but not extra property. Having a phone and an email.
4	Influence of various housing types in comparison between men and women	45-year-old men and women	Married with two children	Medium prestige job, with a medium employment duration, having a lower secondary education	Living in various housing types: an owned house or apartment, a rented apartment, a municipal apartment, or with parents	Owning a car, but not extra property. Having a phone and an email.

5	Influence of car and property ownership between men and women	45-year-old men and women	Married with two children	Medium prestige job, with a medium employment duration, having a lower secondary education	Living in a house or apartment	Having various combinations of ownership of a car or a property. Having a phone and an email.
6	Influence of phone and email ownership between men and women	45-year-old men and women	Married with two children	Medium prestige job, with a medium employment duration, having a lower secondary education	Living in a house or apartment	Owning a car, and extra property. Having various combinations of phone and email ownership.
7	Influence of the education level between men and women	45-year-old men and women	Married with two children	Medium prestige job, with a medium employment duration, having various education levels: academic, higher, incomplete higher, secondary, lower secondary	Living in a house or apartment	Owning a car, and extra property. Having a phone and an email.
8	Verification testing of selected features & profiles					

Source: own elaboration

Appendix 3. Simulation results from the final testing of various age group profiles – decision of approval or denial, and confidence levels*

Trial	Gender	Decision	Confidence	Specific feature: age
Group1	Woman	Denied	0.63	55
	Woman	Denied	0.63	40
	Woman	Denied	0.63	25
	Man	Approved	0.58	55
	Man	Approved	0.58	40
	Man	Approved	0.58	25

Source: own elaboration

*Confidence level indicates the probability of a correctly made decision, on a scale from 0.50 to 1 (applies to all appendices)

Appendix 4. Simulation results from the initial testing of selected occupation status profiles

Trial	Gender	Decision	Confidence	Specific feature: occupation status
4b	Women	Approved	0.53	Retired
	Man	Approved	0.53	
5b	Woman	Denied	0.59	Working
	Man	Denied	0.57	

Source: own elaboration

Appendix 5. Simulation results from the initial testing of selected occupation status profiles

Trial	Gender	Decision	Confidence	**Specific feature:** occupation status
4b	Women	Approved	0.53	Retired (age: old)
	Man	Approved	0.53	
6b	Woman	Denied	0.59	Working (age: young)
	Man	Denied	0.57	

Source: own elaboration

Appendix 6. Simulation results from the final testing of profiles representing various family statuses

Trial	Gender	Decision	Confidence	**Specific feature:** family status
Group2 (1)	Woman	Denied	0.66	Civil marriage
	Woman	Denied	0.64	Married
	Woman	Denied	0.67	Separated
	Woman	Denied	0.71	Single
	Woman	Denied	0.63	Widow
	Man	Approved	0.67	Civil marriage
	Man	Approved	0.68	Married
	Man	Approved	0.63	Separated
	Man	Approved	0.56	Single
	Man	Approved	0.68	Widow

Source: own elaboration

Appendix 7. Simulation results from the initial testing of profiles representing different genders

Trial	Gender	Decision	Confidence	Specific feature
1a	Woman	Denied	0.53	-
	Man	Approved	0.62	-

Source: own elaboration

Appendix 8. Simulation results from the initial testing of profiles representing low-income level

Trial	Gender	Decision	Confidence	Specific feature
12a	Man	Approved	0.51	Low-income level
	Woman	Denied	0.50	Low-income level

Source: own elaboration

Appendix 9. Simulation results from the initial testing of profiles representing family status as single

Trial	Gender	Decision	Confidence	**Specific feature:** family status
4a	Woman	Denied	0.62	Single
	Man	Denied	0.50	Single

Source: own elaboration

Appendix 10. Simulation results from the final testing of profiles representing the number of children

Trial	Gender	Decision	Confidence	**Specific feature:** number of children
Group2 (2)	Woman	Denied	0.64	0
	Woman	Denied	0.62	1
	Woman	Denied	0.62	2
	Woman	Denied	0.62	3
	Woman	Approved	0.55	4
	Man	Approved	0.68	0
	Man	Approved	0.58	1
	Man	Approved	0.56	2
	Man	Approved	0.56	3
	Man	Approved	0.60	4

Source: own elaboration

Appendix 11. Simulation results from the initial testing of profiles representing the number of children

Trial	Gender	Decision	Confidence	**Specific feature:** number of children
11a	Woman	Denied	0.50	0
	Woman	Approved	0.56	1
	Woman	Approved	0.57	2
	Woman	Approved	0.57	3
	Woman	Approved	0.59	4
	Woman	Approved	0.59	5
	Woman	Approved	0.59	6
	Woman	Approved	0.59	7
	Woman	Approved	0.59	8

Source: own elaboration

Appendix 12. Simulation results from the final testing of profiles representing various degrees of prestige of occupation

Trial	Gender	Decision	Confidence	**Specific feature:** prestige of occupation
Group3 (1)	Woman	Denied	0.57	high
	Woman	Denied	0.62	medium
	Woman	Denied	0.64	low
	Man	Denied	0.51	high
	Man	Denied	0.50	medium
	Man	Denied	0.55	low

Source: own elaboration

Appendix 13. Simulation results from the final testing of profiles representing various employment durations

Trial	Gender	Decision	Confidence	**Specific feature:** employment duration
Group3 (2)	Woman	Denied	0.62	long
	Woman	Denied	0.62	medium
	Woman	Denied	0.62	short
	Man	Denied	0.50	long
	Man	Denied	0.50	medium
	Man	Denied	0.50	short

Source: own elaboration

Appendix 14. Simulation results from the initial testing of profiles representing unemployed individuals

Trial	Gender	Decision	Confidence	**Specific feature:** occupation status
7b	**Women**	Denied	0.67	Unemployed
	Man	Denied	0.76	Unemployed

Source: own elaboration

Appendix 15. Simulation results from the final testing of profiles representing various housing types

Trial	Gender	Decision	Confidence	**Specific feature:** housing type
Group4	Woman	Denied	0.65	Co-op apartment
	Woman	Denied	0.62	House / apartment'
	Woman	Denied	0.63	Municipal apartment'
	Woman	Denied	0.66	Office apartment'
	Woman	Denied	0.65	Rented apartment'
	Woman	Denied	0.65	With parents'
	Man	Denied	0.53	Co-op apartment
	Man	Denied	0.50	House / apartment'
	Man	Denied	0.55	Municipal apartment'
	Man	Denied	0.52	Office apartment'
	Man	Denied	0.52	Rented apartment'
	Man	Denied	0.56	With parents'

Source: own elaboration

Appendix 16. Simulation results from the initial testing of profiles representing various housing types

Trial	Gender	Decision	Confidence	**Specific feature:** housing type
1b	**Woman**	Approved	0.56	House / apartment'
	Man	Approved	0.65	
10b	**Women**	Approved	0.57	Co-op apartment
	Man	Approved	0.62	
11b	**Women**	Approved	0.57	Municipal apartment'
	Man	Approved	0.58	
12b	**Women**	Approved	0.51	Office apartment'
	Man	Approved	0.54	
13b	**Women**	Approved	0.55	Rented apartment'
	Man	Approved	0.59	
14b	**Women**	Approved	0.55	With parents'
	Man	Approved	0.56	

Source: own elaboration

Appendix 17. Simulation results from the final testing of profiles representing various combinations of ownership of cars and property

Trial	Gender	Decision	Confidence	**Specific feature:** ownership of cars and property
Group5	Woman	Denied	0.62	Both
	Woman	Denied	0.73	None
	Woman	Denied	0.62	Car
	Woman	Denied	0.72	Property
	Man	Denied	0.54	Both
	Man	Denied	0.65	None
	Man	Denied	0.50	Car
	Man	Denied	0.69	Property

Source: own elaboration

Appendix 18. Simulation results from the final testing of profiles representing various combinations of ownership of phone and email

Trial	Gender	Decision	Confidence	**Specific feature:** ownership of phone and email
Group6	Woman	Denied	0.62	Both
	Woman	Denied	0.67	None
	Woman	Denied	0.64	Phone
	Woman	Denied	0.65	Email
	Man	Denied	0.50	Both
	Man	Denied	0.65	None
	Man	Denied	0.60	Phone
	Man	Denied	0.54	Email

Source: own elaboration

Appendix 19. Simulation results from the final testing of profiles representing various education levels

Trial	Gender	Decision	Confidence	**Specific feature:** education level
Group7	Woman	Denied	0.67	Higher education
	Woman	Denied	0.71	Incomplete higher
	Woman	Denied	0.62	Lower secondary
	Woman	Denied	0.70	Secondary / secondary special
	Man	Denied	0.50	Higher education
	Man	Denied	0.53	Incomplete higher
	Man	Denied	0.50	Lower secondary
	Man	Denied	0.56	Secondary / secondary special

Source: own elaboration

Appendix 20. Simulation results on number of total approved and denied profiles and the confidence scores per each testing phase

Testing phase	Gender	Decision: Approved (count)	Decision: Denied (count)	Total (count)	Average confidence: Approved	Average confidence: Denied	Approval rate
Initial (1)	**Men**	8	1	18	0.58	0.50	89%
	Women	1	8		0.57	0.56	11%
				Average	0.57	0.53	**50%**
Initial (2)	**Men**	9	5	28	0.59	0.71	64%
	Women	9	5		0.56	0.72	64%
				Average	0.57	0.72	**64%**
Final	**Men**	13	27	80	0.60	0.54	33%
	Women	2	38		0.56	0.65	5%
				Average	0.58	0.59	**19%**

Source: own elaboration

Appendix 21. Summary of simulation results on approved and denied profiles in total

Gender	Total Approved			Total Denied			Approval rate
	Count	Percentage (column)	Percentage (row)	Count	Percentage (column)	Percentage (row)	
Men	30	71%	48%	33	39%	52%	**48%**
Women	12	29%	19%	51	61%	81%	**19%**
Both	42	100%	33%	84	100%	67%	**33%**

Source: own elaboration

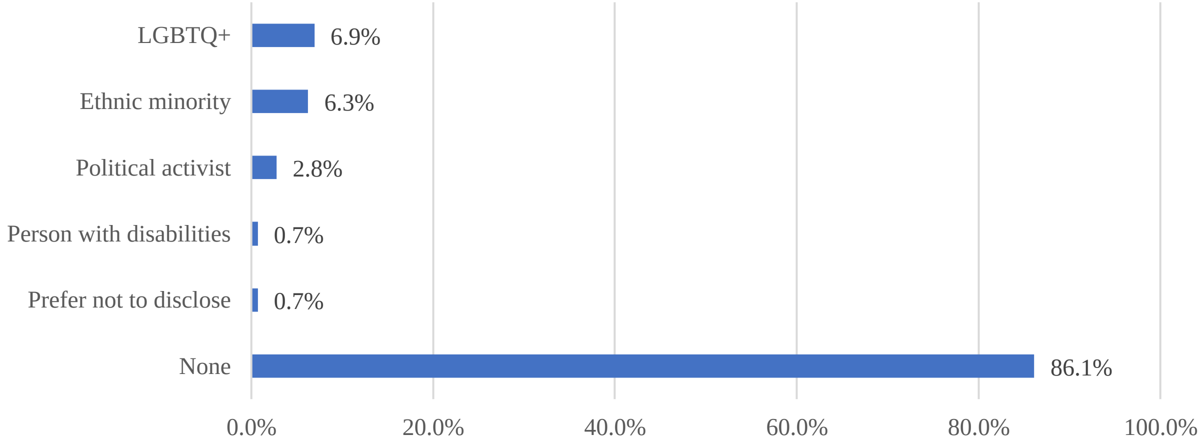
Appendix 22. Socio-demographic profiles of survey respondents (n=144, in percent)

Age intervals	**Percent**	**Cumulative Percent**
18-21	7.6	7.6
22-25	43.8	51.4
26-29	5.6	56.9
30-33	4.9	61.8
34-37	6.3	68.1
38-41	9.7	77.8
42+	22.2	100
Gender	**Percent**	**Cumulative Percent**
Women	72.2	72.2
Men	27.8	100
Education level	**Percent**	**Cumulative Percent**
Doctor	3.5	3.5
Master	22.9	26.4
Bachelor	45.1	71.5
Secondary	27.8	99.3
Below secondary	0.7	100.0

Community	Percent	Cumulative Percent
The capital city	38.2	38.2
Suburban area of the capital city	14.6	52.8
Another large city	26.4	79.2
Small city or town	12.5	91.7
Rural area	4.2	95.8
Abroad	4.2	100.0

Source: own elaboration

Appendix 23. Representation of potentially vulnerable groups towards algorithmic bias among the survey respondents (n=144, in percent)



Source: own elaboration

Appendix 24. Descriptive statistics on general attitudes of respondents towards technologies and AI on a scale from 1 to 5* (n=144)

Measure	Statements					
	1.1. Impact of technology on society is positive.	1.2. I am increasingly encountering artificial intelligence (AI) in my daily life.	1.3. AI algorithms can make decisions more efficiently (quicker, with less errors, more streamlined) than humans.	1.4. I fully trust decisions made by AI systems.	1.5. AI algorithms can have bias towards certain groups of people.	1.6. How confident are you on the previous answer?
N	144	144	144	144	144	144
Mode	4.000	4.000	4.000	3.000	4.000	5.000
Median	4.000	4.000	3.000	2.000	3.000	4.000
Mean	3.556	3.757	3.368	2.486	3.252	3.806
Std. Deviation	0.867	1.117	0.937	1.044	1.146	1.202
Minimum	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	5.000	5.000	5.000	5.000	5.000	5.000

Source: own elaboration

* Scale from 1 to 5, where one indicates strongly disagree and five indicates strongly agree.

Appendix 25. Chi-squared test results on the views of respondents on the positive impact of technology based on socio-demographic characteristics (n=144)

Socio-demographic features		Value	df	p value
Gender	X²	8.2	4	0.085
	N	144	-	-
Age	X²	26.83	24	0.312
	N	144	-	-
Education	X²	30.03	16	0.018
	N	144	-	-
Community	X²	30.18	20	0.067
	N	144	-	-
Vulnerable groups	X²	12.9	4	0.012
	N	144	-	-

Source: own elaboration

Appendix 26. Breakdown of respondent evaluation on impact of technology based on their affiliation with a potentially vulnerable group (n=144, in percent)

1.1. Impact of technology on society is positive	Affiliation with a vulnerable group (%)		Total (%)
	No	Yes	
Strongly disagree	0.8	5.0	**1.4**
Disagree	8.1	5.0	**7.6**
Neither agree, nor disagree	41.1	20.0	**38.2**
Agree	40.3	35.0	**39.6**
Strongly agree	9.7	35.0	**13.2**
Total:	**100**	**100**	**100**

Source: own elaboration

Appendix 27. Chi-squared test results on the respondent evaluation of their trust in decision made by AI based on socio-demographic characteristics (n=144)

Socio-demographic features		Value	df	p value
Gender	X²	4.711	4	0.318
	N	144	-	-
Age	X²	16.15	24	0.882
	N	144	-	-
Education	X²	19.099	16	0.264
	N	144	-	-
Community	X²	15.474	20	0.749
	N	144	-	-
Vulnerable groups	X²	7.091	4	0.131
	N	144	-	-

Source: own elaboration

Appendix 28. Chi-squared test results on the respondent evaluation of AI algorithm potential for bias based on socio-demographic characteristics (n=144)

Socio-demographic features		Value	df	p value
Gender	X^2	2.079	4	0.721
	N	144	-	-
Age	X^2	26.514	24	0.328
	N	144	-	-
Education	X^2	10.315	16	0.850
	N	144	-	-
Community	X^2	21.696	20	0.357
	N	144	-	-
Vulnerable groups	X^2	3.274	4	0.513
	N	144	-	-

Source: own elaboration

Appendix 29. Contingency table of respondent views on AI bias towards certain groups of people and their confidence levels* (n=144, in percent)

1.5. AI algorithms can have bias towards certain groups of people	1.6. How confident are you on the previous answer? (%)					Total
	1	2	3	4	5	
Strongly disagree	0.00	0.0	0.0	3.5	4.3	**7.8**
Disagree	0.00	1.7	0.9	7.0	8.7	**18.3**
Neither agree, nor disagree	0.00	2.6	8.7	11.3	6.1	**28.7**
Agree	0.87	6.1	6.1	11.3	7.0	**31.3**
Strongly agree	0.00	0.0	0.0	3.5	10.4	**13.9**
Total:	**0.87**	**10.4**	**15.7**	**36.5**	**36.5**	**100.0**

Source: own elaboration

* Scale from 1 to 5, where one indicates not being confident at all and five indicates being very confident

Chi-Squared Test

	Value	df	p value
X^2	31.085	16	0.013
N	115		

Source: own elaboration

Appendix 30. Breakdown of credit application histories of respondents and the application results (n=144, in percent)

2. Have you ever applied for a credit?	Percentages	3. Have you ever been denied credit application? (%) Yes	Not applicable	No	Total
Yes	Of total	18.8	0	24.3	**43.1**
	Within row	43.5	0	56.5	**100**
No	Of total	0	56.9	0	**56.9**
	Within row	0	100	0	**100**
Total:		**18.8**	**56.9**	**24.3**	**100**

Source: own elaboration

Appendix 31. Breakdown of credit application histories of respondents among different age groups (n=144, in percent)

2. Have you ever applied for a credit?	Percentages	Age groups (%) 18-21	22-25	26-29	30-33	34-37	38-41	42 or older	Total
Yes	Of total	1.4	6.3	2.8	3.5	4.9	8.3	16.0	**43.1**
	Within row	3.2	14.5	6.5	8.1	11.3	19.4	37.1	**100**
No	Of total	6.3	37.5	2.8	1.4	1.4	1.4	6.3	**56.9**
	Within row	11.0	65.9	4.9	2.4	2.4	2.4	11.0	**100**
Total:		**7.6**	**43.8**	**5.6**	**4.9**	**6.3**	**9.7**	**22.2**	**100**

Source: own elaboration

Chi-Squared Test

	Value	df	p value
X^2	52.157	6	< .001
N	144		

Source: own elaboration

Appendix 32. Breakdown of credit application results of respondents among different age groups (n=144, in percent)

3. Have you ever been denied credit application?	Percentages	Age groups (%)							Total
		18-21	22-25	26-29	30-33	34-37	38-41	42 or older	
Yes	Of total	0.7	2.1	1.4	2.8	2.8	3.5	5.6	**18.8**
	Within column	9.1	4.8	25.0	57.1	44.4	35.7	25.0	**18.8**
Not applicable	Of total	6.3	37.5	2.8	1.4	1.4	1.4	6.3	**56.9**
	Within column	81.8	85.7	50.0	28.6	22.2	14.3	28.1	**56.9**
No	Of total	0.7	4.2	1.4	0.7	2.1	4.9	10.4	**24.3**
	Within column	9.1	9.5	25.0	14.3	33.3	50.0	46.9	**24.3**
Total:	Of total	7.6	43.8	5.6	4.9	6.3	9.7	22.2	**100**
	Within column	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**-**

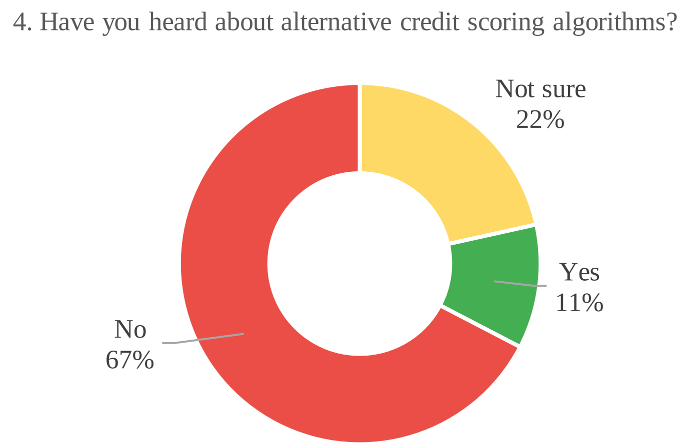
Source: own elaboration

Chi-Squared Test

	Value	df	p value
X^2	59.045	12	< .001
N	144		

Source: own elaboration

Appendix 33. Awareness of the respondents regarding alternative credit scoring algorithms (n=144, in percent)



Source: own elaboration

Appendix 34. Chi-squared test results on the possibility of respondent participation in alternative creditworthiness assessment based on socio-demographic characteristics (n=144)

Socio-demographic features		Value	df	p value
Gender	X^2	2.43	4	0.66
	N	144	-	-
Age	X^2	32.78	24	0.11
	N	144	-	-
Education	X^2	21.50	16	0.16
	N	144	-	-
Community	X^2	14.5	20	0.80
	N	144	-	-

Source: own elaboration

Appendix 35. Breakdown of respondent participation in alternative creditworthiness assessment based on their views on AI decisions as more efficient (n=144, in percent)

5. Would you incline to be evaluated under alternative creditworthiness evaluation?	Percentages	1.3. AI algorithms can make decisions more efficiently (quicker, with less errors, more streamlined) than humans (%)					Total (%)
		Strongly disagree	Disagree	Neither agree, nor disagree	Agree	Strongly agree	
No	Within row	15.0	25.0	30.0	25.0	5.0	**100**
	Of total	2.1	3.5	4.2	3.5	0.7	**14.0**
Rather no	Within row	3.1	9.4	43.8	34.4	9.4	**100**
	Of total	0.7	2.1	9.8	7.7	2.1	**22.4**
Maybe	Within row	0.0	25.0	43.8	21.9	9.4	**100**
	Of total	0.0	5.6	9.8	4.9	2.1	**22.4**
Rather yes	Within row	2.0	4.0	28.0	58.0	8.0	**100**
	Of total	0.7	1.4	9.8	20.3	2.8	**35.0**
Yes	Within row	0.0	0.0	55.6	22.2	22.2	**100**
	Of total	0.0	0.0	3.5	1.4	1.4	**6.3**
Total:		**3.5**	**12.6**	**37.1**	**37.8**	**9.1**	**100**

Source: own elaboration

Chi-Squared Tests

	Value	df	p value
X^2	32.249	16	0.005
N	144		

Source: own elaboration

Appendix 36. Main reasons identified by respondents who do not agree to be evaluated by alternative credit scoring system, based on the given scenario (n=144, in percent)

5.2. Main reasons for not agreeing for alternative credit scoring evaluation	**Percentage**
Fear of discrimination	13%
Distrust of AI and its decisions	29%
Distrust in the method / not fully understanding its working principles	44%
Unwillingness to share the necessary information	62%
Fear of the information leaks	4%
Total:	**100%**

Source: own elaboration

Appendix 37. Descriptive statistics – central tendencies of respondent evaluations, on a scale from 1 to 5*, for statements based on the offered scenario (n=144)

Measure	**Statements**				
	6.1. Alternative credit scoring evaluations offer improved precision of results	6.2. Explanation of principles or basis of decision is unnecessary	6.3. There is no difference between AI and human evaluations	6.4. This method will make fairer decisions and traditionally discriminated groups will benefit	6.5. Higher trust in AI systems of banks than other financial institutions
Mode	3.0	1.0	2.0	4.0	3.0
Median	3.0	2.0	2.0	3.0	3.5
Mean	3.1	2.4	2.5	3.2	3.5
Std. Deviation	1.0	1.3	1.1	0.9	0.9
Minimum	1.0	1.0	1.0	1.0	1.0
Maximum	5.0	5.0	5.0	5.0	5.0

Source: own elaboration

* Scale from 1 to 5, where one indicates strongly disagree and five indicates strongly agree.

Appendix 38. Descriptive statistics of respondent evaluations for statements based on the offered scenario (n=144, in percent)

Evaluation	Statements (%)				
	6.1. Alternative credit scoring evaluations offer improved precision of results	6.2. Explanation of principles or basis of decision is unnecessary	6.3. There is no difference between AI and human evaluations	6.4. This method will make fairer decisions and traditionally discriminated groups will benefit	6.5. Higher trust in AI systems of banks than other financial institutions
Strongly disagree	6.9	29.9	22.2	2.1	1.4
Disagree	15.3	27.1	28.5	21.5	9.0
Neither agree, nor disagree	45.8	17.4	28.5	25.7	31.9
Agree	25.7	20.8	16.0	34.0	31.9
Strongly agree	6.3	4.9	4.9	3.5	10.4
N/A	0.0	0.0	0.0	13.2	15.3

Source: own elaboration

Appendix 39. Pearson's correlation between answers of respondents regarding improved precision of results offered by alternative credit scoring and their personal contact with AI in everyday life (n=144)

Pearson's Correlations

Variable		1.2. I am increasingly encountering artificial intelligence (AI) in my daily life.
6.1. Alternative credit scoring evaluations offer improved precision of results	Pearson's r	0.234
	p-value	0.005

Source: own elaboration

Appendix 40. Chi-squared test results on the evaluation of fairer decision-making of alternative credit scoring based on socio-demographic characteristics of respondents (n=144)

Socio-demographic features	Value		df	p value
Gender	X^2	0.237	4	0.994
	N	144	-	-
Age	X^2	21.026	24	0.637
	N	144	-	-
Education	X^2	21.169	16	0.172
	N	144	-	-
Community	X^2	25.495	20	0.183
	N	144	-	-
Vulnerable groups	X^2	1.744	4	0.783
	N	144	-	-

Source: own elaboration

Appendix 41. Chi-squared test results on the evaluation of willingness to share selected data types by respondents representing vulnerable groups (n=144)

Data types used for credit scoring evaluations		Value	df	p value
Demographic	X^2	2.877	4	0.579
	N	144	-	-
Utility payment	X^2	2.511	4	0.643
	N	144	-	-
Telco	X^2	4.825	4	0.306
	N	144	-	-
Geolocation	X^2	3.508	4	0.477
	N	144	-	-
Health	X^2	1.9	4	0.754
	N	144	-	-
Social media	X^2	1.427	4	0.839
	N	144	-	-
Used app and web history	X^2	2.659	4	0.616
	N	144	-	-

Source: own elaboration

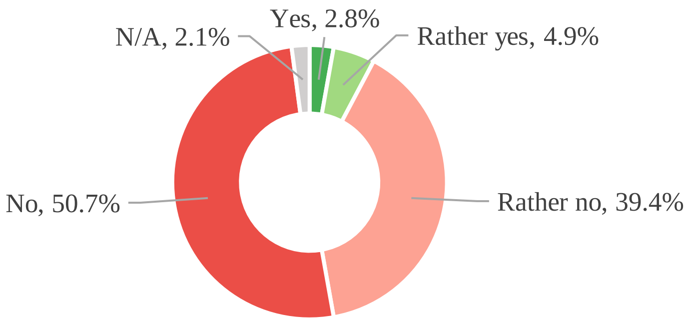
Appendix 42. Descriptive statistics – central tendencies of respondent evaluations, on a scale from 1 to 5*, for readiness to share various types of information for credit scoring purposes (n=144)

Measure	Demographic data	Telco data	Used apps and web history	Social media insights	Geolocation	Utility payment data	Health data
N/A	5	3	1	1	2	3	10
Mode	4	3	2	1	4	4	2
Median	4	3	2	2	3	3	2
Mean	3.78	2.89	2.20	2.20	2.84	3.17	2.40
Std. Deviation	1.21	1.26	1.09	1.16	1.35	1.31	1.27
Minimum	1	1	1	1	1	1	1
Maximum	5	5	5	5	5	5	5

Source: own elaboration

* Scale from 1 to 5, where one indicates that the respondents would highly unlikely share the information and five indicates that the respondent would highly likely share it.

Appendix 43. Answers of respondents regarding their willingness to install tracking software for credit scoring purposes (n=144)



Source: own elaboration

Appendix 44. Descriptive statistics – central tendencies of respondent evaluations, on a scale from 1 to 5*, regarding their comfort levels of sharing sensitive personal information with financial institutions (n=144)

Measure	Sexual orientation	Gender identity	Religious affiliation	Political ideology	Membership in associations
N/A	6	5	4	6	5
Mode	1	3	3	1	3
Median	3	3	3	2	3
Mean	2.77	3.25	2.98	2.39	3.32
Std. Deviation	1.51	1.47	1.46	1.26	1.36
Minimum	1	1	1	1	1
Maximum	5	5	5	5	5

Source: own elaboration

* Scale from 1 to 5, where one indicates completely uncomfortable and five indicates completely comfortable

Appendix 45. Chi-squared test results on fear of information leaks based on socio-demographic characteristics of respondents (n=144)

Socio-demographic features	Value		df	p value
Gender	X^2	4.679	1	0.031
	N	144	-	-
Age	X^2	8.671	6	0.193
	N	144	-	-
Education	X^2	7.192	4	0.126
	N	144	-	-
Community	X^2	10.286	5	0.068
	N	144	-	-
Vulnerable groups	X^2	0.737	1	0.391
	N	144	-	-

Source: own elaboration

Appendix 46. Fear of information leaks between men and women respondents (n=144, in percent)

Fear of information leaks	Gender (%)		Total (%)
	Women	Men	
No	22.1	40.0	**27.1**
Yes	77.9	60.0	**72.9**
Total:	**100**	**100**	**100**

Source: own elaboration

Appendix 47. Chi-squared test results on fear of discrimination based on socio-demographic characteristics of respondents (n=144)

Socio-demographic features		Value	df	p value
Gender	X^2	0.213	1	0.644
	N	144	-	-
Age	X^2	9.646	6	0.140
	N	144	-	-
Education	X^2	4.211	4	0.378
	N	144	-	-
Community	X^2	7.309	5	0.199
	N	144	-	-
Vulnerable groups	X^2	3.434	1	0.064
	N	144	-	-

Source: own elaboration

Appendix 48. Chi-squared test results on participation in alternative credit scoring if it is proven to have bias based on socio-demographic characteristics of respondents (n=144)

Socio-demographic features		Value	df	p value
Gender	X^2	7.538	4	0.110
	N	78	-	-
Age	X^2	20.533	24	0.666
	N	78	-	-
Education	X^2	8.793	12	0.720
	N	78	-	-
Community	X^2	14.066	20	0.827
	N	78	-	-
Vulnerable groups	X^2	4.582	4	0.333
	N	78	-	-

Source: own elaboration

Appendix 49. Chi-squared test results on participation in alternative credit scoring if it is proven to have bias free based on socio-demographic characteristics of respondents (n=144)

Socio-demographic features		Value	df	p value
Gender	X^2	5.902	4	0.207
	N	65	-	-
Age	X^2	39.542	24	0.024
	N	65	-	-
Education	X^2	15.575	12	0.211
	N	65	-	-
Community	X^2	34.222	20	0.025
	N	65	-	-
Vulnerable groups	X^2	3.837	4	0.429
	N	65	-	-

Source: own elaboration

Appendix 50. Respondent willingness to be evaluated under alternative credit scoring if it is proven to be bias free based on the community type (n=144, in percent)

10.b Would you like to be evaluated under alternative credit scoring, if it was proven that it is bias free?	Community (%)						Total (%)
	Another large city	Rural area	Small city or town	Suburban area of the capital city	The capital city	Abroad	
No	5.9	40.0	7.7	10.0	0.0	0.0	7.7
Rather no	29.4	20.0	23.1	20.0	20.0	20.0	23.1
I do not know	23.5	0.0	30.8	10.0	13.3	0.0	16.9
Rather yes	35.3	0.0	38.5	60.0	66.7	80.0	47.7
Yes	5.9	40.0	0.0	0.0	0.0	0.0	4.6
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Source: own elaboration

Appendix 51. Respondent willingness to be evaluated under alternative credit scoring if it is proven to be bias free based on the age group (n=144, in percent)

10.b Would you like to be evaluated under alternative credit scoring, if it was proven that it is bias free?	Age (%)							Total (%)
	18-21	22-25	26-29	30-33	34-37	38-41	42 or older	
No	0	0.0	0.0	0.0	16.7	20.0	16.7	**7.7**
Rather no	28.6	16.0	100.0	0.0	16.7	0.0	33.3	**23.1**
I do not know	42.9	24.0	0.0	50.0	0.0	0.0	5.6	**16.9**
Rather yes	28.6	60.0	0.0	50.0	33.3	80.0	38.9	**47.7**
Yes	0	0.0	0.0	0.0	33.3	0.0	5.6	**4.6**
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Source: own elaboration