University of Nottingham
UK | CHINA | MALAYSIA

A Consumer Centred Investigation of Differentially Private Risk Assessment Models in Consumer Credit

Thesis submitted to the University of Nottingham for the degree of **Doctor of Philosophy, 13th November 2024.**

Ana Rita Pena

14252585

Supervised by

Derek McAuley
Simon Preston
Alexandra Lang
Katie Severn
Madeline Hallewell
Steve Benford

Signature _____

Date _____ / _____ / _____

Abstract

Differential Privacy (DP) is a technology which allows one to gather aggregate information without compromising individual privacy. Over the last few years, it has become the state-of-the-art privacy-enhancing technology. DP has been implemented by several Big Tech companies, as well as governmental bodies, but research in applied contexts is still at a very early stage. As differential private algorithms have an inherent accuracy-privacy trade-off and no guarantees of an equal accuracy loss for different dataset subgroups, when applied in practical settings the accuracy drop could have significant impacts to consumers.

This thesis aims to understand the social and technical repercussions of implementing DP in Credit Risk Assessment Models in the UK Consumer Credit Industry from a consumer centred perspective. To achieve this, a sociotechnical approach was employed using a combination of qualitative and technical work. The first qualitative studies were an exploratory user interviews about the application process and an interview-based industry stakeholder consultation. The technical element was the implementation and comparison of different differentially private decision tree-based algorithms. The thesis culminates in an interactive game study to gather consumers' attitudes towards the implementation.

Findings from the technical study found that the DP algorithms had a neg-

ligible accuracy drop for specific amounts of privacy when compared to a non-private algorithm and rare occasions of disparate accuracy loss. Triangulating these findings with the knowledge on the workings of the consumer credit industry from the industry consultation we can deduce that if DP was implemented, the majority of consumers would not be significantly affected, with the exception of the consumers that are closer to the threshold of being denied credit.

The implementation of DP would be dependent on the amount of accuracy loss and regulatory encouragement, according to the industry consultation findings. To compensate for the implementation, lenders could change their credit policy to account for the small increase in uncertainty in the risk scores. This could make credit less accessible, which goes against regulatory aims, and hence not likely to have regulatory support.

Consumers also had very mixed views regarding the implementation of DP, as they would rather have better financial options than protect their personal data. These findings are based on the interactive game study, which communicated potential scenarios of the implementation of DP in the risk assessment model in the loan application process to gather consumers' attitudes towards the technology.

Based on these findings DP is unlikely to be implemented, as lenders would require some regulatory encouragement which seems unlikely unless there is a shift in public opinion. This work contributes to the underrepresented area on usable DP and consumers' requirements and attitudes towards the loan application process in the UK consumer credit industry.

Acknowledgements

I would first like to thank my supervision team: Alexandra Lang, Katie Severn, Madeline Hallewell, Simon Preston and Derek McAuley. Thank you for supporting and guiding me in this project, even when I struggled to communicate how it all came together. I have grown and learned much from you in these last five years.

To everyone in the Horizon CDT, thank you for creating such a supportive and creative environment. I would further like to thank my Industry Partner, Capital One UK.

Thank you to everybody who joined my research, for your energy and time, it goes without saying that without you this work wouldn't be possible.

Finally, in some personal notes, Mel, Lena and Luis and Teri, without you I wouldn't have made it, thank you for the encouragement, the laughs, and lifting me when needed. Thank you to all my friends and family, with a special note of gratitude to my Horizon 2019 cohort friends and colleagues for the companionship.

This thesis is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (UKRI Grant No. EP/S023305/1).

Contents

List of Tables

List of Figures

Abbreviations

DP Differential Privacy.

ML Machine Learning.

PET Privacy Enhancing Technologies.

DP-SGD Differentially Private Stochastic Gradient Descent.

CRA Credit Reference Agency.

APR Annual Percentage Rate.

FCA Financial Conduct Authority.

BoE Bank of England.

PRA Prudential Regulation Authority.

PATE Private Aggregation of Teacher Ensembles.

DAL Disparate Accuracy Loss.

ODAL Opposite Disparate Accuracy Loss.

DT Decision Tree.

GBM Gradient Boosting Machine.

LDP Local Differential Privacy.

NPV Net Present Value.

ROC Receiver Operating Characteristic.

AUC Area Under ROC Curve.

LR Differentially Private Logistic Regression.

SRF Smooth Random Forest.

DPGBDT Differentially Private Gradient Boosting Decision Tree.

Mathematical Notation

D**:** dataset.

X**:** set of characteristics of each data instance of D.

$\hat{Y} = f(X)$**:** predicted outcome.

ε**:** privacy parameter/privacy budget of DP.

δ**:** probability of failure of DP.

$\mathcal{M}$**:** randomized algorithm/mechanism/query.

$\Delta\mathcal{M}$**:** sensitivity of $\mathcal{M}$ (maximum amount of change in a query's output if one instance of the dataset is changed).

$S^*(\mathcal{M}, D)$**:** smooth sensitivity (relaxation of sensitivity based on the actual dataset used).

Chapter 1

Introduction

Over the last couple of decades, there has been an increase in the variety and quantity of Machine Learning (ML) applications in a broad range of areas including finance [78], healthcare [5], and social networking [6], among many others. Currently, most industries and services have to some degree implemented tools based on ML algorithms [37]. The increased implementation of ML is both used as a reason and also only made possible due to the collection of large amounts of personal data, which is further supported by the increase in computational power at an accessible price to be able to store and analyse this data [77]. This change in the technological paradigm has led to the development of new business models based on this technology and the associated data [170, 182].

ML is an area of study focused on computational systems that can learn from data, and identify patterns with little human intervention, using different algorithms and statistical models. Usually, the output of a ML algorithm is a trained model, i.e. a function that can accurately perform the task it has been trained for. Due to the workings of ML, trained models can accidentally leak information about some of its data points,

a phenomenon called 'memorization'. Notwithstanding data leaks on the part of the models, there are also different types of data attacks: privacy attacks, where attackers can extract information from individuals used in the training dataset, and security attacks, attacks where attackers gain unauthorized access to a system and possibly release data to the public [171]. To prevent such attacks a variety of technologies have been created, Differential Privacy (DP) is one of those technologies which is considered state-of-the-art in preventing privacy attacks.

The work presented in this thesis aims to investigate the potential future effects of implementing this technology within the context of the UK consumer credit industry, specifically in the risk assessment model of the loan application process.

Figure 1.1 shows a summary of the loan application steps.

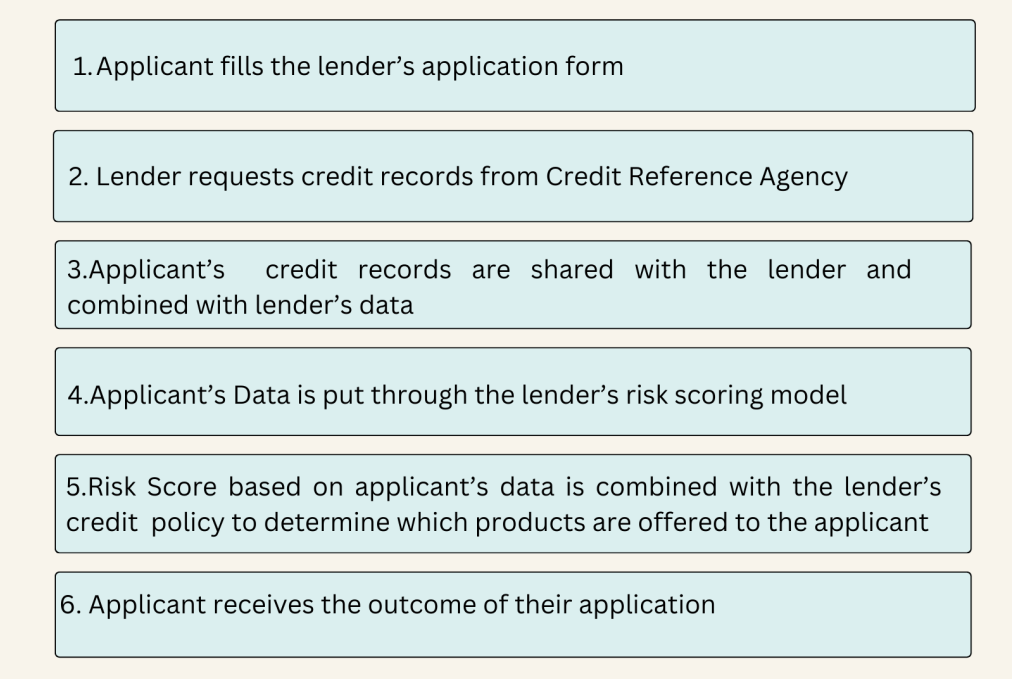


Figure 1.1: Summary of the Loan Application Process. Based on the findings of the Industry consultation study presented in Chapter 5.

As seen in Figure 1.1, there are several stakeholders involved in a simple

loan application process: the consumer, who applies and provides their personal information, the credit reference agency (CRA) which provides data to the lender, and the lender which uses the consumer's personal data to make a decision over if the applicant will be given access to credit. The decision process involves two main components: the risk assessment model - which score the applicants probability of defaulting on payment and the credit policy which accounts for the external economic risk and the potential profitability of the loan given to the consumer.

If the risk assessment model element of a consumer credit loan application was changed to a differentially private risk model, what would change in the industry and how would this implementation of DP ultimately affect consumers. This is the question I explore in my thesis, in other words, I investigate the potential impact of a differentially private risk scoring model in the loan application process.

DP is a privacy enhancing technology which has an associated privacy-accuracy trade-off. In this thesis we will investigate how different privacy-accuracy behaviours would affect consumers if DP is implemented within this context. Below is a short description of how DP works.

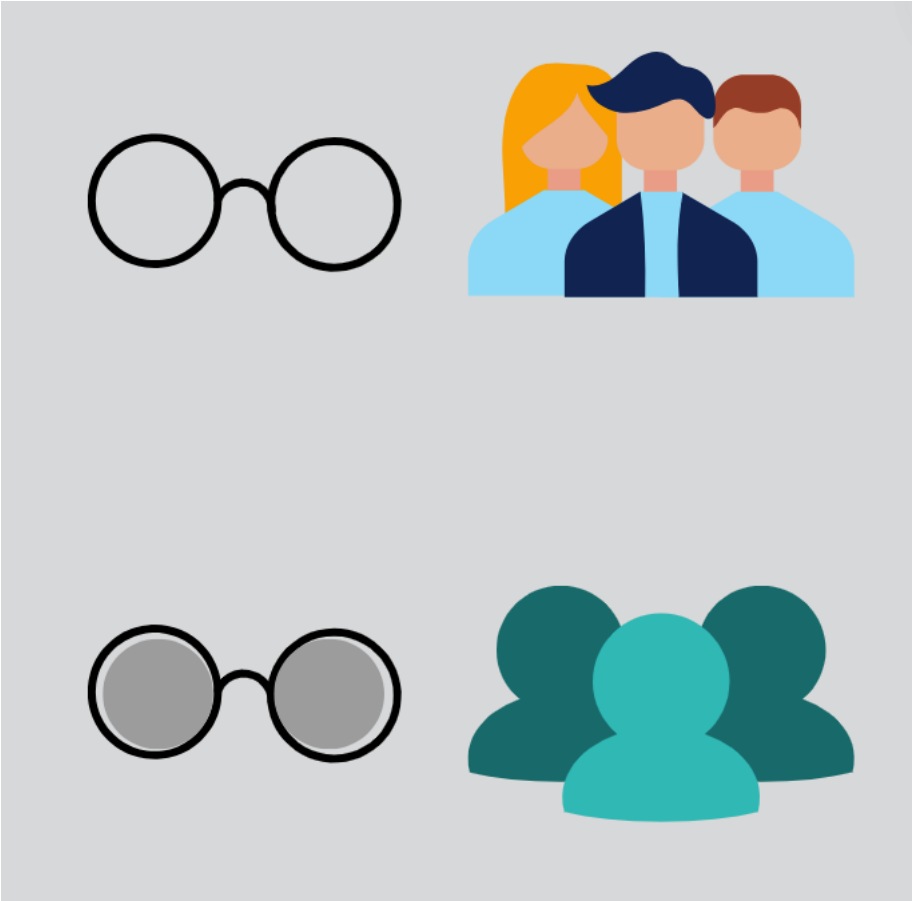


Figure 1.2: Visual analogy of workings of DP

DP works like looking at a group of people (representing the dataset) through some glasses (representing the model). If there is no fog in the glasses (represents noise and in this context privacy parameter) we can get the number of people in the group (aggregate information) and distinguish between different people even if they are very similar, however, if the glasses start fogging up then we can still know the number of people but can no longer differentiate between people as they start to just look like blurs the more there is fog (to represent the accuracy privacy trade-off). This technology will be described in more detail after the Research Questions, in section 1.2.

1.1 Research Questions

The overall Research Question I aim to answer in this work is:

- **RQ:** What are the repercussions to customers of the implementation of DP in Credit Risk Assessment Models in UK consumer credit industry applications?

Here repercussions focuses mainly on the consumer. However, to fully understand the impact on the consumer it is essential to understand the repercussions on the industry and how that is passed on to the consumer. My approach to this research question considers all stakeholders involved while still having a consumer -centred approach, as these are the most affected and least powerful stakeholders.

The design of this research question is underpinned on the principles of responsible research and innovation (RRI) and ethics, i.e. understanding the potential consequences of implementing this technology and using those findings to inform the decision-making of technology design and implementation.

The work in this thesis is set within the context of the UK's credit industry. This is an industry that processes large amounts of sensitive personal data through statistical and ML models, generally opaque to the consumer and has an ubiquitous impact on consumer's lives through the proxy of their financial situations. As such, it is essential to understand the potential repercussions of technology implementation within this industry.

In order to start answering the overall research question above a series of sub-research questions were defined.

As this thesis is based on the consumer's perspective the first sub research question is of an exploratory nature and tries to capture their experiences with the industry and consequent consumers' perceptions.

- **RQ1:** What are consumers' attitudes regarding current loan application practices?

This question was addressed in an exploratory semi-structured interview study.

The second sub-research question elicits from the perspective of the industry, i.e. it provides the necessary contextual knowledge to which build on the rest of the questions and research activities.

- **RQ2:** What are the processes of the consumer loan application that impact outcomes?

This question was created as when reviewing the literature the processes and decision making behind this industry remained quite opaque, more detail in Chapter 2.

While RQ1 and RQ2 provided essential contextual information, RQ3 and RQ4 start focusing specifically on the technology I am studying in this work. RQ3 is focused on the perspective of the Industry.

- **RQ3:** What are the UK consumer credit industry perspectives on DP implementation in the risk assessment model of the loan application?

This questions was created to help address and understand what would happen to the industry processes if DP was implemented (based on a variety of hypothetical scenarios).

Both RQ2 and RQ3 were answered in the same research activity via an online semi-structured interview with industry stakeholders. The choice to combine two research questions in one activity was based on optimising the output from participants once they were involved and avoid having to overcome barriers to recruitment in this sensitive industry.

RQ4 focuses on the technology itself, and explores the intrinsic privacy-accuracy trade-off associated with DP:

- **RQ4:** What is the accuracy drop behaviour for DP Decision Tree based models applied to credit risk assessment models?

As it will be seen in the next section and Chapter not all combinations of different differentially private algorithms and datasets have the same privacy-accuracy trade-off behaviour. As within this work there is a specific application and industry in which the research is set, it is important to understand the technology in that specific setting.

This research question was answered by implementing a variety of DP Decision Tree based models (as these are commonly used in the industry, based on the literature review and findings from RQ2) with financial datasets to understand the different privacy-accuracy trade-offs.

As stated in the beginning of this section the work in this thesis takes a consumer centred approach, as such the last sub research question brings the new knowledge back to the consumer.

- **RQ5:** What are consumers' attitudes towards DP implementation in the loan application process?

This is addressed in a focus group based around a interactive game board

activity, designed to represent the loan application process and diverse DP behaviours.

1.2 Introduction to Differential Privacy

Privacy Enhancing Technologies (PET) are a series of technologies whose aim is to help preserve the privacy of individuals, with different technologies being more or less adequate for different scenarios. Differential privacy (DP) is the technology I will focus on in this work, has become a state-of-the-art PET for the private release of statistical information [123]. It provides a mathematical guarantee of privacy independently of the attacker's computational power and auxiliary data, the properties which make it stand out from other PETs. It guarantees that given a study or query its results will not change considerably if any individual takes part or not, i.e. if a single row of data is added, or no new data is added, DP guarantees that the outcome will be similar in the two cases. It allows us to gather general information about the population without compromising individual's privacy. DP can be achieved in three main ways:

- input perturbation: by using differentially private input
- output perturbation: adding noise to the non-private output
- in-learning perturbation: by making changes to the learning algorithm so that the perturbation happens within learning

Figure 1.3 showcases a simple example of how DP works, which will be built upon further in Chapter 2. It exemplifies how adding noise to the output (answer) of an average height query for two similar datasets makes it impossible to calculate the height of the missing individual from the

second dataset. This is possible where no noise is added as seen in Figure 1.3. Compared with other PETs such as k-anonymity [153] ,DP has

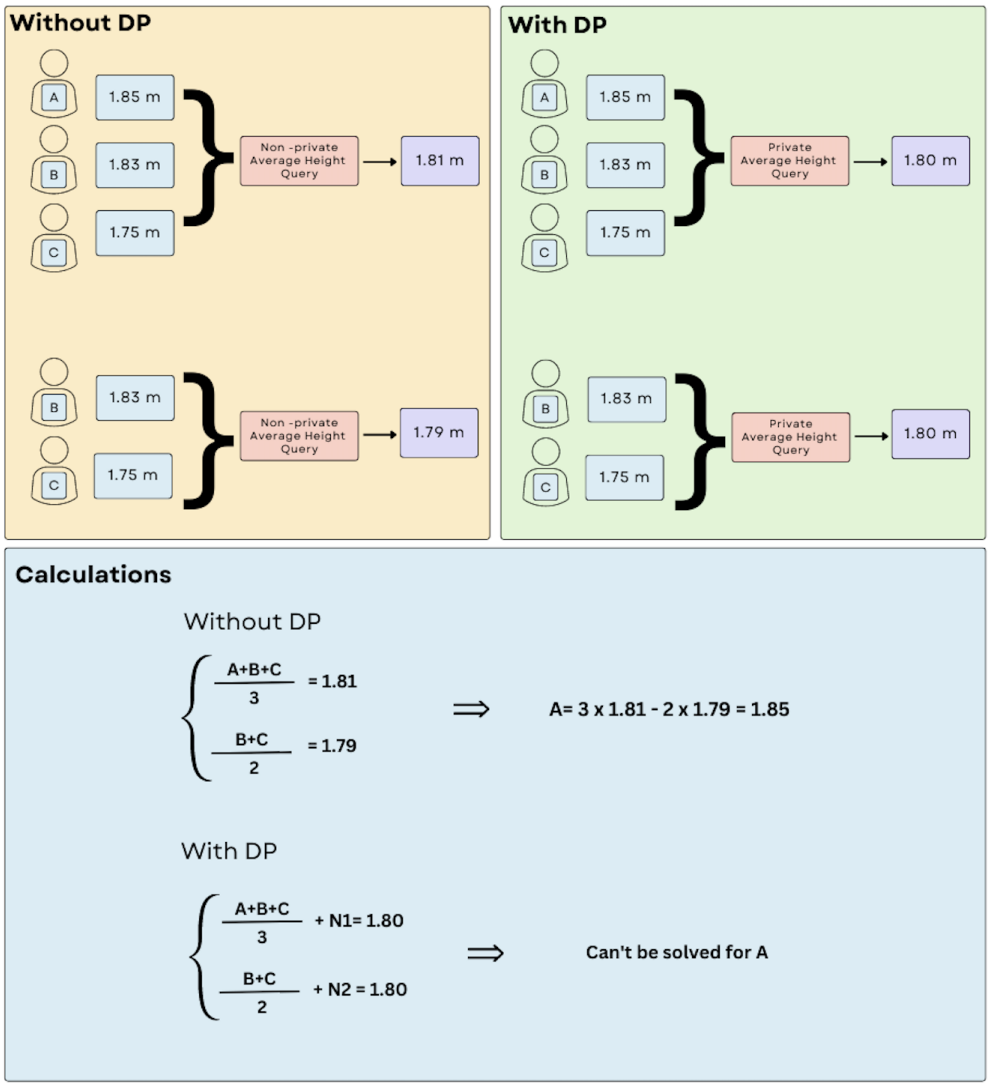


Figure 1.3: Simple example of DP's working

stronger privacy guarantees, due to being a mathematical definition. DP has been implemented by a variety of big tech companies such as Google, Apple, Microsoft and even by the US Census Bureau, partly because it is a mathematical guarantee [175].

However, this uniformisation of what privacy is, through DP becoming the state-of-the-art PET and being based on technical mathematical concepts, dismisses the contextual nuances of users' privacy requirements. Due to its

technical language DP is hard to understand in practical terms for non-specialists, making its implementation easily performative, by for example, companies reporting that they are using DP, but having the "DP parameter" set so low it has no practical impact [144]. On the more positive side, DP can be a useful tool in efforts to increase transparency of algorithms due to some of its properties, as DP allows one to have third-party queries onto a model while still maintaining privacy, allowing for an explanation of outcomes and different explainability metrics.

DP comes with an associated privacy-accuracy trade-off, due to the addition of the noise. This trade-off is dependent on the privacy level required, which can be set by the privacy budget parameter. The higher the privacy budget, the less strict privacy is and hence less noise is added. As a consequence accuracy remains high. On the other hand, the smaller the privacy budget, the stricter privacy is and consequently more noise is added leading to a lower accuracy. Figure 1.4 shows an example of the privacy-accuracy trade-off behaviour. In the case limits: no privacy which means high privacy budget (left side of Figure 1.4.) and infinite privacy with a privacy budget of 0 (right side of Figure 1.4.).

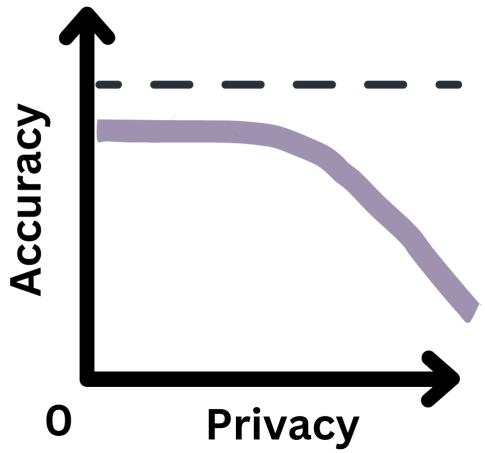


Figure 1.4: Simple example of privacy-accuracy trade-off

In the limit behaviour where we have no privacy the model's accuracy

should be the same as a non-private model and on the other hand, if we have infinite privacy then the model will be allocating outcomes randomly so the accuracy will showcase this (in the case of only two outcomes with the same amount of each the accuracy would be 0.5).

Some differentially private algorithms, specifically Differentially Private Stochastic Gradient Descent (DP-SGD), have shown to have a disparate accuracy loss for training dataset subgroups. Some of the studies [15] found the disparate accuracy loss happened in underrepresented and complex training subgroups, and others that it also occurred in majority training subgroups [95], where the disparate accuracy drop was dependent on the gradient distributions of each subgroup [176]. As DP is now implemented in a variety of applied "real-world" scenarios, disparate accuracy drops for some subgroups of the training datasets could lead to significant impacts for all stakeholders involved.

As DP is deployed in applied contexts the corresponding social institutions shape its implementation and DP in turn shapes the social institutions (see more in Chapters 2 and 3). It is therefore important to study DP from a sociotechnical perspective (acknowledging and accounting for the mutual shaping of the social dynamics of the industries implementing DP and DP itself) in different contexts. This research area focusing on a sociotechnical perspective towards DP is currently in its very early development stages.

My work aids the understanding and evaluation of DP in applied contexts, specifically focusing on the impact of DP in the Risk Assessment Models from a user impact perspective. These models are part of the loan application processes in the UK Consumer Credit Industry. Historically technology has played an important and formative role within this industry in its recent past and present, which will be highlighted in Chapter

2.1. My work further addresses the understudied area of usable DP (see Chapter 2.2.3.), focusing on capturing users'/consumers' attitudes towards this technology and its behaviours.

1.3 Research Environment

My PhD is of an interdisciplinary nature and is part of the Horizon CDT (Center for Doctoral Training) program. The opportunity for the PhD came with a partnership with an external institution, Capital One UK. As Capital One is a credit card company it grounded my work within the consumer credit industry.

As my work looks into DP from a sociotechnical perspective, it draws elements from computer science as well as the social sciences, which will be highlighted in more detail in Chapter 3. Academically my work is grounded in the emergent research area of Usable DP. Usable DP is a research area within DP, which focuses on its usability and how to best communicate DP with a range of stakeholders, being itself an interdisciplinary area.

I started my PhD in October of 2019 just months before the start of the COVID-19 pandemic, as this fell within the initial stage of the PhD I was able to plan my research activities to follow social distancing and lock-down restrictions. This meant that the initial exploratory study with the consumer (Chapter 4) was completed online, as well as the Industry Consultation (Chapter 5). At this point, COVID restrictions began to ease but I found that running the study online allowed me to have a bigger pool of potential participants. During the course of my PhD, COVID-19 had a significant financial impact on people's lives and in the broader UK economy [12]. The ensuing Cost of Living Crisis which saw an increase

in the number of people unable to afford their bills [92, 86], discussed in more depth in Chapter 3.1.4. . This economic context undoubtedly shapes consumers' attitudes towards the Credit Industry.

1.4 Contributions

The research activities presented in this document contribute to the Usable DP research area. The entirety of the thesis is one of the first studies of DP from a sociotechnical perspective in an applied industry (with only one previous paper published by Aslan et al in the Healthcare context [8]). I argue that to fully realise the impact of this technology it is necessary to study it in context and not just focus on its technical performance.

Figure 1.5 summarises the main contributions based on the findings of this thesis.

The order of the contributions in the table below is in order of readiness of application in the real world.

Thesis Contributions (non-extensive)	
1. New ways to communicate DP	
Summary: Design of the glasses analogy and interactive game to communicate DP **Who is it useful to:** • people implementing DP (range of industries and government departments) • DP academics (especially those working within Usable DP) • consumer - better understanding of what happens to their data with DP	**Findings in:** • Chapter 1.1 • Chapter 7.2
2. New data that shows consumer desire for more transparency in the credit industry	
Summary: Participants express wanting more information regarding how their data is used in the application process, more details regarding the process and explanations of outcomes. **Who is it useful to:** • Credit industry - could lead to improvement of consumer relationships • Regulator (FCA)	**Findings in:** • Chapter 4.3 • Chapter 7.3
3. New knowledge on the privacy-accuracy trade-off of DP Decision Tree models	
Summary: DP- Gradient Boosting Decision Tree and Smooth Random Forest models have comparable performances to non-private models and no significant disparate accuracy loss for data subgroups. **Who is it useful to:** • people implementing DP (range of industries and government departments) • DP academics	**Findings in:** • Chapter 6.3
4. New knowledge on consumer attitudes towards DP (and its varying privacy accuracy trade-off behaviours) in the consumer credit industry	
Summary: Varied views on prefered DP behaviour and random element associated with DP is seen as unfair in the loan application setting. **Who is it useful to:** • Credit industry - could lead to improvement of consumer relationships • Regulator (FCA) • DP academics (especially those working within Usable DP)	**Findings in:** • Chapter 7.3
5. New knowledge on the potential impact of DP implementation on the credit industry	
Summary: DP would probably not negatively impact applicants disparately, however access to credit might become more restricted due to changes to credit policy **Who is it useful to:** • Credit Industry • Regulator • Consumers • DP academics	**Findings in:** • Chapter 5.3 • Chapter 6.3 • Chapter 7.3

Figure 1.5: Thesis Contributions (non-exhaustive)

These contributions will be discussed in more detail in Chapter 8.

1.5 Thesis Structure

Figure 1.6 shows the different thesis chapters and their interconnections.

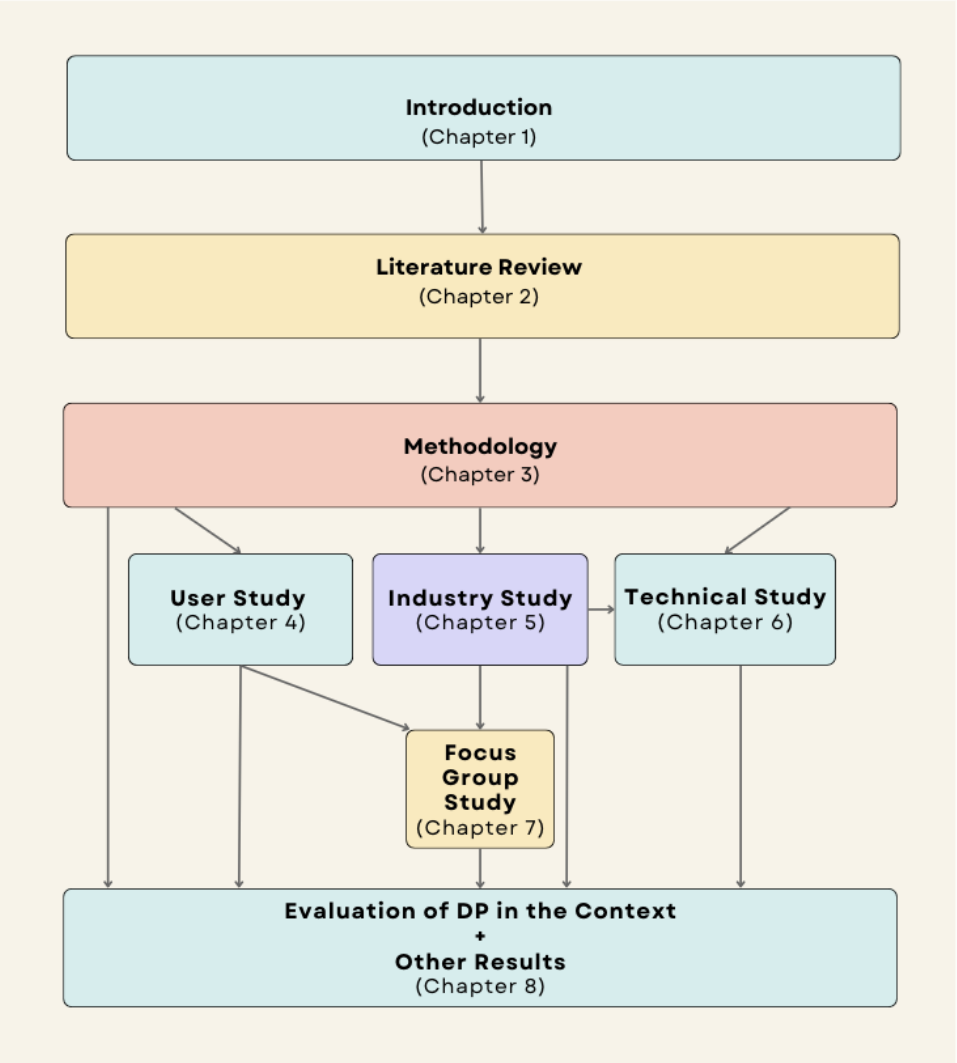


Figure 1.6: Thesis Outline

Chapter 2: Literature Review This chapter gives us the background for the research studies and it is divided into three main sections:

- History and associated technologies of the Consumer Credit Industry in the UK
- Differential Privacy and Usable DP

- Consumers' sensemaking and perceptions of algorithms

Chapter 3: Methodology

Summarises different research philosophical positions and associated methodologies. I state my philosophical orientation and discuss how it has shaped the design of the different research activities and justify the methodologies chosen for each of them.

Chapter 4: Attitudes and Experiences with Loan Applications - Consumer Study

This chapter describes the first research study, an interview-based study designed to empathise and understand consumers' sensemaking of their experiences when applying for loans, as well as their attitudes regarding automation, data sharing and fairness of the process (**RQ1**). The study consists of a semi-structured interview and a post-interview survey. The interview included topics such as participants' previous experiences with the loan application process, data used, and the automation and fairness of the process.

Chapter 5: UK Consumer Credit Industry Consultation

This interview-based study with participants who work or have worked within or with the UK Consumer Credit Industry and was designed to ground informal knowledge of the workings of the consumer credit industry (gained through the course of an internship) on participants' data. The interview was divided into two parts. The first was designed to better understand of the Consumer Credit Ecosystem, including gaining a better awareness of the role of the different stakeholders, and interactions between them. This section of the study also focused on understanding the process

of new tech implementation in the industry: which stakeholders are involved and how? Which external factors are at play? (**RQ2**). The second part of the interview focused on understanding the importance and current practices regarding privacy in the industry. Furthermore, it was also designed to gather stakeholders' attitudes towards DP and the potential impacts of its implementation in the industry (**RQ3**).

Chapter 6: Differentially Private Decision Tree -based Models -Technical Study

The Differentially Private Decision Tree-based Model study is of an exploratory nature and consists of the implementation of different DP models on three credit-related open-source datasets to compare each algorithm's effect on accuracy and subgroup accuracy (**RQ4**). A Smooth Random Forest and different configurations of a Differentially Private Gradient Boosting Machine (DPGBDT) were trained with three different datasets and compared to a differentially private logistic regression and a non-private GBM (using the library LightGBM).

Chapter 7: Differentially Private Consumer Credit Imaginaries - Gamified Focus Group Study

The study consists of a group game-based interactive activity designed to understand how users/consumers perceive the implementation of Differential Privacy in different scenarios. This creative approach was developed to educate focus group participants in DP, in a more interactive, accesible and inclusive manner. The activity was piloted with two different groups to guarantee the efficacy of the communication. The study involves an in-person focus group with a game board style activity. This game then provided the structure for a focus group activity and in-depth discussion on the scenarios created, DP (**RQ5**) and participants' attitudes towards

the Industry.

Chapter 8: Discussion

This chapter starts by summarising the findings which address each of the research questions, building up to answer the general research question: What are the repercussions to customers of the implementation of DP in Credit Risk Assessment Models in UK consumer credit industry applications?

The rest of the chapter addresses both the limitations of the work and contributions as well as outlining the directions for future work.

Chapter 2

Literature Review

This thesis aims to understand the impact on consumers of the potential implementation of Differentially Private Risk Assessment models in consumer credit loan applications. As a starting point to design research inquiries to help us answers the overall research question it is necessary to understand the context in which the thesis is based on.

The first section of this chapter gives a short historical overview of the UK consumer credit industry, so that we can understand how it came to be what it is today and then goes on to explore the industry today, especially focusing on the role of technology. The second element that is necessary to understand is the technology itself, Differential Privacy, which is the focus of the second section of the chapter. The initial part defines what it is, its properties and commonly used mechanism summarised in a simple example to help better understand the mathematical concepts. The section goes on to explore different privacy-accuracy trade-offs found in the literature and the novel area of Usable DP the closest body of literature to the work present in this thesis. As the thesis focuses on the impact to consumer the third section of this chapter, similarly to the previous

Usable DP section, focuses on consumers/general public general attitudes, perceptions and sensemaking of algorithms. This last section is important to design Chapter 4's research inquiry and to interpreter the data collected. While the consumers' sensemaking of algorithms literature is not directly related to the thesis, as most studies in this literature are based on scenarios where participants deal with algorithms directly that is not the case in the loan application scenario. However, the literature is still important to understand how consumers generally view technology of this type.

As discussed in Chapter 1 the work presented in this thesis is a social science approach to the technology DP. It further takes a sociotechnical approach (which will be discussed in more detail in the following chapter). Therefore, needing to understand:

- the interactions between the social institutions and networks in which the technology might be deployed in
- the way these institutions make decisions regarding technology (literature in the section 2.1.)
- the behaviour of the technology itself (literature in section 2.2.) within the context.

The literature presented in this chapter is a prerequisite to combine this knowledge with the research activity findings and understand the potential repercussions on consumers of the implementation of DP in the risk assessment model of the loan application.

2.1 Financial Sector and the Consumer Credit Industry

Throughout history agreement of future payments in exchange for the supply of goods has been commonly used in day-to-day exchanges, and these exchanges were largely based on a relationship of trust between both parties:

> "From the tab in the bar, to the slate in the grocer's shop, small amounts of credit provided on the basis of established relationships have allowed people to get by and get the things they need. Credit makes economies work and has a social purpose. " [172]

The initial commercialisation of consumer credit started in the post-war period, with the first consumer credit cards being created in the late 1950s in the US and being introduced in the UK in 1966 (the first card was Barclaycard introduced by Visa) [2].

The following Figure 2.1 summarises the UK consumer credit industry from the 70's until today.

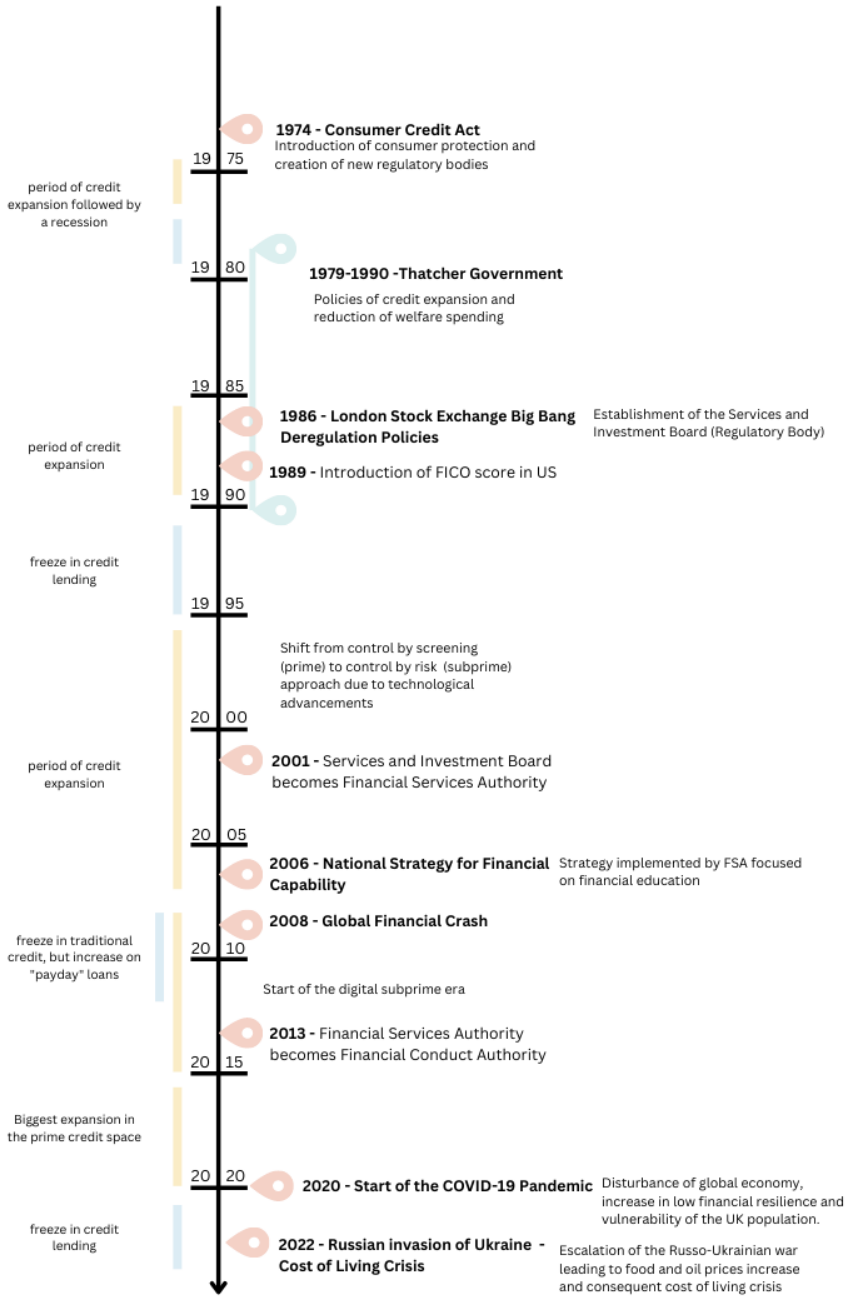


Figure 2.1: UK Consumer Credit Industry Timeline

The timeline shows the periodical boom and bust of the credit industry on the left side, and on the right hand side it highlights important changes and events in the UK and global economy as well as events related to changes of regulation and technology implementation.

A particularly influential time for the credit industry and the country in

general was the period of Thatcher's administration. This was characterised by notions of self-reliance leading to policies of credit expansion, deregulation of industries and the reduction of the welfare state, which for some, was substituted with credit products to cover basic necessities. The implementation of statistical scores highly impact the consumer credit industry, which will be discussed in the next subsection.

2.1.1 The rise of the Subprime: the start of modern lending

Starting in 1994 there was an increase in lending which only ended with the global crash of 2008. This increase in lending was highly influenced by different changes in the industry: automation of credit scoring, risk-based pricing and securitisation [77].

Due to the technological and computational gains of the '80s and '90s Credit Reference Agency (CRA) could now store, analyse and share data more easily and at a bigger volume, which at that time was mainly negative data [77].

Up until this point approaches to loan allocation were based on rules and thresholds that the applicants needed to fulfil based on their credit score. This is known as a credit control-by-screening approach. As a result of these strict lending rules, only a small amount of people had access to credit if they had a good score as the aim was to minimise risk. This neglected the rest of the population entirely [130]. The most commonly used metric was the FICO score (Fair Isaac Corporation score) which was introduced in the consumer credit industry in 1989.

The FICO score became widespread and standardised due to its imple-

mentation by a series of US Government bodies. With the existence of a common metric to standardise and compare loan products, debt could be bought and traded as an investment product, a process called securitization. Securitization allows banks to have the capital to further expand their credit products [130].

The initial automation of credit scoring was based on the control-by-selection approach. Sets of rules that were automated, which brought speed and consistency but could not predict default.

It was by the mid-90s that rule-based decisions started to give space to statistical scoring, which could predict the probability of default based on the data held by the CRA's and statistical techniques.

This led to the creation of risk-based pricing, where scores and the consequent segmentation of the population according to credit risk were used to determine the price of credit consumers would be offered to offset said financial risk. In turn, this resulted in an expansion of credit as previously financially excluded people could now gain access to credit [130]. The two major segments of credit products then became:

- **Prime:** which consists of applicants with a low probability of default and consequently products with low APR (Annual Percentage Rate)
- **Subprime:** which consists of applicants with a higher probability of default and associated products with high APRs

> "Empirically derived credit scoring techniques have created a new kind of consumer whose calculability defied conventional assumptions about the binary nature of creditworthiness. [The credit score] became a platform for creative design work that

> brought lines of risk-calibrated products, both mortgages and securities, into existence." [130]

The combination of all these changes and innovations led to a rapid credit expansion. The market quickly became saturated, which saw borrowers competing for customers, specifically defaulting customers as these were a source of profit [77].

In 2006 a National Strategy for Financial Capability was implemented by the FSA (Financial Services Authority, the precedent of the current Financial Conduct Authority (FCA)), which focused on financial education. This demarcated a turn in the approach of the FSA turning the failures of the markets into an individual problem of the consumer.

2.1.2 Post 2008 Global Financial Crisis and the Digital Subprime

The credit expansion period mentioned in the previous section burst with the 2008 global financial crash. The crash was attributed to the increase in unsustainable unsecured subprime mortgages and their purchase as investment products by global financial institutions. Once the real estate market burst in the US, the value of these products collapsed and destabilised the world's economy [40].

Within the specific UK context, there have been several changes to the industry as a result of the 2008 Financial Crisis and recent technological evolution. One impact of the financial crisis was a decrease in consumers' trust in the UK Banking system. This led to a regulatory shift to prioritise the service of the consumer over sales [4]. In the last decade UK banks

and building societies have been working on regaining consumer trust by improving their services, increasing transparency and making banking more accessible [4].

In the UK regulatory field, the Irresponsible Lending Guidance was published in 2008 [122], which states that lenders should evaluate the affordability of loans to the borrowers to advert potential negative impacts on the lives of the borrowers. This addition of the affordability check takes into consideration the loan product terms the applicants are applying for and is like a control-by-screening approach.

Reform of the industry happened through collaboration between the regulators and the industry, which measures included greater transparency to consumers, and providing services to allow customers to switch current accounts more easily. Furthermore, in 2012 the Financial Services Act established the FCA and its acting powers. Currently, the FCA is still the institution in charge of regulating the Consumer Credit Market in the UK.

Between 2009 and 2010, 35,000 people were declared insolvent every three months and by the end of 2010, the Consumer Credit Counselling Service had over 110000 people on Debt Management Plans [77], showcasing the broad and devastating impact of the crisis.

In the period following the financial crisis, there was a rapid growth of short-term credit products, commonly known as 'payday loans' [74]. These products were sought as households were forced to find money to cover living costs, in a period of high unemployment.

These types of credit products are characterised by very heavy Annual Percentage Rate (APR), with some reaching 5000% APR. However, in 2015 a cap was introduced on the amount of interest on payday loans, decreasing

it to 1500% APR [51].

Furthermore, with the wider spread of mobiles and computing devices and internet access over this time period, certain small start-up businesses explored non-financial data for credit scoring of short-term high-interest loan products, a practice and tendency named digital subprime by Deville [51].

> "While being broadly concerned with the redefinition of conventional credit scoring practices for online lending, the term [digital subprime] encapsulates three more specific interlinked tendencies. First, offering loans that are short-term and high cost – what are sometimes referred to as payday loans - with a customer base assumed to have poor or non-existent credit histories. Second, the reengineering of forms of online social connectivity and influence-based assessment to determine creditworthiness. And third, the reengineering of forms of data mining and algorithmic analysis. In this sense, the digital subprime can be seen as transferring to the field of credit scoring sets of 'big data' techniques and logics most commonly associated with online marketing (...). " [51]

The techniques used in digital subprime collect information on applicants based on their online behaviour via cookies, such as location data, e-commerce shopping habits, and online footprint, and combine this with more traditional financial data to be able to access creditworthiness. Using these alternative data sources allows companies to offer credit to a subgroup of the population that previously did not have any, due to their non-existent credit history [51].

One example of a digital subprime company from the UK, that employed

such methods is Wonga. Wonga was a subprime lender that had very high-interest rates and became bankrupt as a result of several controversies, such as the case of a highly indebted teenager committing suicide after a bank account clearance [74], and the change in the regulation which put a cap on interest rates [51]. Deville [51] exposed that the Wonga website used nudging techniques on their two sliders, which controlled the amount of the loan and the repayment time. Nudging techniques change the environment, in this case, the initial position of the sliders to elicit an action like applying for a loan. Wonga, among other similar creditors, showed interest in using data collected from Facebook (before the Cambridge Analytica controversy), however, Deville had no confirmation that this data was collected and used. The example of the Wonga case was discussed here as it is still in memory in the British collective memory and was mentioned by participants of different studies.

2.1.3 Recent Picture

Most of the growth in consumer credit in 2018 was in the prime sector, consumers who are less likely to suffer from financial distress, with only a small increase in the amount of subprime consumers. At the start of 2020, due to the Covid-19 pandemic, the UK entered the first of several periods of lockdown where many employees were furloughed and the economy was highly impacted. The resulting lockdowns had an enormous impact on people's lives as showcased by the Financial Lives Survey results.

In February 2021, the FCA published the latest version of their Financial Lives survey with over 16,000 participants [12]. The survey found that from 2017 to February 2020 the percentage of consumers who were vulnerable (meaning suffering from poor health, a traumatic life event, low financial

resilience or low capability) decreased from 51% to 46%. This decrease was caused by fewer people being digitally excluded and fewer having low financial resilience. However, in October 2020 this number had increased to 53% due to the Covid-19 pandemic. Financial resilience, which is not being over-indebted and being able to withstand financial shock, follows a similar trend. People who were less capable of coping with financial shock were mainly unemployed adults, renters, adults with a household income of less than £15,000 and Black adults.

The Covid-19 pandemic did not impact everyone equally, as a third of households in the UK were able to repay debt and borrow less due to the reduced expenditure caused by lockdowns. Over the same time period, there was an increase in approximately 60% of people running out of money before the end of the week and month, where lower-income households were twice as likely as high-income households to have increased their use of consumer credit during lockdown [172]. The previous statistics highlight the role that the Covid-19 pandemic had in increasing inequality.

Both the industry and the FCA reacted quickly to the lockdown. High-cost and short-term credit products were severely reduced however these types of products have since gone back to pre-pandemic levels. The FCA created a set of guidance for consumer credit and mortgages that included allowing payment deferrals during lockdowns as well as masking credit files, so that consumers did not suffer long-term consequences for an unprecedented financial event outside of their control.

The cost of living started to increase at the start of 2022 and inflation reached a 41-year high in October 2022 at 11.1% [86]. Food and energy prices (domestic gas prices increased by 129% and domestic electricity prices by 66%) have risen sharply in this period, partly caused by Rus-

sia's invasion of Ukraine. This has led to an increase in the cost of living, especially for low-income households who spend a larger proportion than average on food and energy [86].

As people's living costs increase, the amount of consumer credit borrowed also increases. From May to June 2022 an additional 1.8 billion pounds in loans have accrued, as people struggle to afford their bills and hence resort to credit [92]. To combat inflation the Bank of England has been raising interest rates to loan products [86].

2.1.4 AI and ML in the Financial Industry

With the closure of branches over the pandemic, the role of online banking and its related technology in the sector increased and was more clearly understood. According to a report by the Bank of England, the number of financial services that use ML continues to grow and is expected to continue, where credit is the second sector where ML is critical to the business, where treasury takes the leading place [13].

The majority of firms that employ these technologies do it within their current governance, as the deployment is not seen as risky. However, some concerns remain related to the possibility of bias representativeness and the lack of interpretability of these models. In terms of the types of models used the most common are Decision Tree based models, however, the choice of model probably is related to the specific application [13].

At the time of the report (2022), ML in credit was mainly used to supplement existing scorecards or as part of the pre-approval process. These ML applications process unstructured data and/or large volumes of data. However, credit decisioning of personal data based on ML is not widespread

across respondents [19]. In order to address the changes and potential risks and benefits which arise with the widespread implementation of ML techniques in different types of applications within the financial sector, the Bank of England (BoE) alongside the FCA and the Prudential Regulation Authority (PRA) have published a Discussion Paper to consider the views of all relevant stakeholders on the use of AI in the sector. The aim of the paper was to understand if there is a need for regulatory changes, which shape this would take and clarifying the way in which current regulations apply to different factors of ML [19].

Furthermore, in June 2022 the Government shared their plans to reform the Consumer Credit Act which have come into force in 2023 [161]. This showcases that the industry is constantly changing and adapting both by part of firms and the regulators.

Overall, the credit industry has transformed significantly in the time of its existence both by the implementation of new technologies, new approaches to pricing (e.g. risk-based pricing) and control, as seen in the Figure 2.2. There is now a lot wider access to credit by the general public compared to the times of Prime lending, however at what cost?

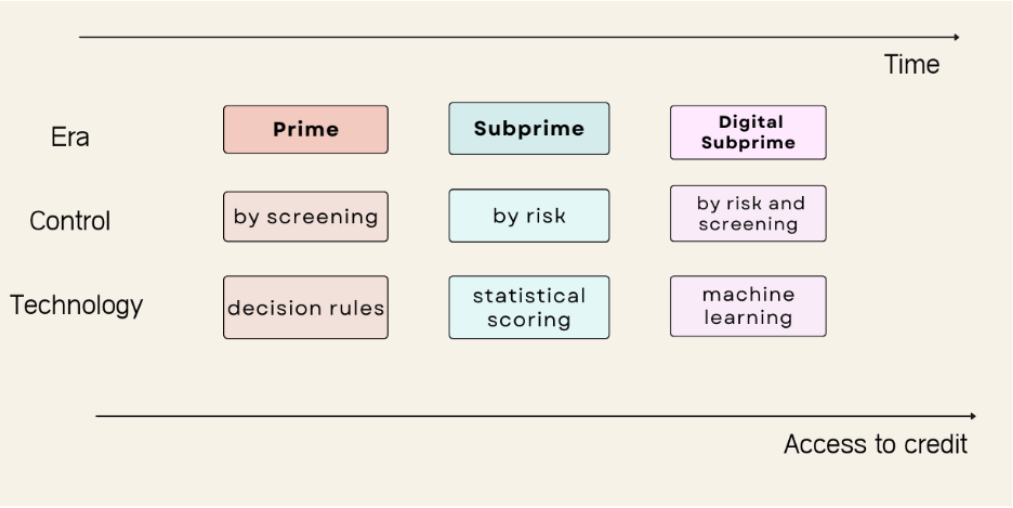


Figure 2.2: Credit Industry Eras

> "The harms which can arise in sub-prime credit markets must be balanced against the harm of not accessing credit at all." [172]

Presently, some lower-income households have to use credit products to be able to cover their living costs, due to the small role that the welfare state plays, ending up paying high amounts in interest due to their higher risk. There are already some firms in the space of sustainable and lower cost credit for those in the subprime markets, such as Fair 4 All Finance, however, these sorts of enterprises need to be expanded as called by the Woolard Review [172]. The industry will continue to change and impact consumers' lives in a way that long-term effects cannot be fully understood yet. This is evident with the rise of 'Buy Now, Pay Later' products which have largely evaded the FCA regulatory umbrella.

2.1.5 Consumer's attitudes towards the Industry

The development and implementation of technology in the industry has changed the way consumers access, interact and transact from online and mobile banking to instant loans and easier access to credit to mobile wallets [34]. From the perspective of the lender, digital technology is seen as an opportunity to gain a competitive advantage and new avenues for profit [165, 34].

Research on the effect of digitalisation on consumer credit has mainly focused on the perspective of the industry, with little evidence gathered on the consumer perspectives [34]. Ironfield-Smith et al. [93] report the results of a financial wellbeing survey from 2003 to gather attitudes towards consumer credit. It found consumer attitudes towards credit were gen-

erally positive despite general uncertainty regarding how loan limits are calculated. With the implementation of increasingly complex technologies over the last two decades the decision-making process has become even more opaque. The FCA's report "Financial Lives 2020 survey: the impact of coronavirus" [12] gathered consumers' attitudes towards the overall financial industry. Trust and confidence in the industry has increased since 2017 but still remains low. Banks are some of the most trusted institutions, but there has been a decrease in trust in credit card companies. Consumer trust may have been eroded by problematic interactions with these services including, poor customer services, IT system failure and unexpected fees and charges [12]. Whilst limited, the previous work in this area highlights that there is an opportunity to improve service design and the consumer experience of credit applications, which could ultimately contribute to improved trust in the sector, as well as the already mentioned more systemic changes to the industry.

2.2 Differential Privacy

DP is a mathematical definition (Definition 1), if an algorithm is differentially private (it upholds definition 1) it guarantees that given a study or query its results will not change considerably if any individual takes part or not. In other words, DP is not a specific type of technology but a mathematical definition, if any algorithm fulfils the conditions set in Definition 1, then its behaviour will have specific properties (which will be discussed in more detail briefly) and one can be sure that if the dataset fed into this algorithm changes very slightly (change in one data point) the results will be so close that it will be impossible for a third party to infer about which dataset was used and consequently about specific data points membership

in the dataset.

DP allows one to gather general information (benefiting those setting the query due to their ability to gather information) about the population without compromising an individual's privacy (benefiting the individuals who are part of the dataset as there is no way to prove if they are part of the dataset or not). This definition is particularly appealing as it is independent of the adversary's computational power, i.e. person or organisation trying to gather information which they should not have access to using a series of computational tools. DP has some useful properties such as: resilience to post-processing, composition and group privacy.[58]. These properties will be discussed in more detail in the following paragraphs.

Differential Privacy can be achieved in three main ways: by using differentially private input (perturbing the input), by perturbing the output of an algorithm and finally by making changes to the learning algorithm so that the perturbation happens within learning, e.g. perturbing the gradient in stochastic gradient descent. However, due to the addition of noise (i.e. perturbation) DP inherently comes with a privacy-accuracy trade-off.

Throughout the chapter we will assume the following notation: $\hat{Y}$ is the predicted outcome for a single data point, X, which takes the form $\hat{Y} = f(X)$. The set of all data instances is D. Furthermore, we take Y to be the true outcome . If we have a binary classification task $\hat{Y} \in \{0, 1\}$, a multi-class classification $\hat{Y} \in \{0, 1, ..., k\}$ where k is the total number of classes and finally $\hat{Y} \in \mathbb{R}$ for regression tasks, while in the context of probabilities of default it would be $\hat{Y} \in [0, 1]$.

Definition 1. ((ε, δ)-Differential Privacy):

A randomized algorithm $\mathcal{M}$ is (ε, δ)-differentially private if for all $\mathcal{S} \subseteq$

Range($\mathcal{M}$) and for all datasets D, D' such that $\| D - D' \|_1 \leq 1$, i.e. D, D' are neighbouring datasets only differing in one data point:

$$\Pr[\mathcal{M}(\text{D}) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[\mathcal{M}(\text{D'}) \in \mathcal{S}] + \delta,$$

where the probability space is over the outcomes of mechanism $\mathcal{M}$ [59].

In other words, the above definition means that for any two neighbouring datasets, i.e., datasets that only differ in one entry (same as one person), the results of a query (or algorithm $\mathcal{M}$) for the two datasets are bounded by an exponential factor of ε, and with a probability of failure given by δ. **If a differentially private algorithm is applied to neighbouring datasets the changes in the results will be very small, where the level of change is dependent on the parameter ε, and it will then be impossible for a third party to infer about which dataset was used and consequently about specific data points membership in the dataset.** Both these parameters are set by the person implementing the query $\mathcal{M}$. **The smaller the values of ε and δ the better the privacy guarantee.**

Within the context of training a ML model, the mechanism $\mathcal{M}$ corresponds to the model's training algorithm. At the end of training the output will be a (ε, δ)-differentially private model $\hat{Y} = f(X)$.

2.2.1 DP Properties

The appeal of the DP guarantee stems from its fundamental properties which are defined below.

Definition 2. (Resilience to post-processing):
If a ε-DP mechanism gives us output $f(X)$, then the output of any func-

tion performed on $f(X)$ is also ε-differentially private. This means that the outcome $f(X)$ can be reused without creating any more privacy loss, furthermore at times the post-processing can decrease ε and therefore improve privacy [125].

Definition 3. (Sequential Composition):
Suppose we have two private mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$ with ε_1 and ε_2, if we perform them sequentially the final mechanism $\mathcal{M}$ is $(\varepsilon_1 + \varepsilon_2)$-differentially private[58]. This means that if we perform two tasks on datasets with points in common, the privacy budget will add up (and hence if considering a fixed total privacy budget we will allocate a smaller budget for each task).

Definition 4. (Parallel Composition):
Suppose we have two private mechanisms M_1 and M_2 with ε_1 and ε_2,if we apply them on disjoint datasets the overall mechanism M will be $max\{\varepsilon_1, \varepsilon_2\}$-differentially private[58].This means that if we perform two tasks on datasets with no points in common, the privacy budget will simply be the biggest out of the two (and therefore if considering a fixed total privacy budget we will allocate a larger budget for each task).

2.2.2 How to achieve privacy?

The way most mechanisms satisfy the privacy definition is by introducing noise drawn from random probability distributions. The amount of noise added to guarantee ε-differential privacy is dependent on the sensitivity of the mechanism.

For example, if our mechanism calculates the length of the training dataset, if we add a record the maximum possible change in the output is 1 (the sensitivity), but if our mechanism outputs the maximum value in the training dataset, then the maximum possible change to the output by adding a record is infinite. In more formal terms the sensitivity of a mechanism gives an upper bound on how much we must perturb the mechanism to preserve privacy, by calculating how much noise we would need to add in the worst case scenario of all possible neighbouring datasets [59].

Definition 5. (Sensitivity):
The sensitivity of a randomized algorithm $\mathcal{M}$ for all D, D' such that $\| D - D' \|_1 \leq 1$ is:

$$\Delta\mathcal{M} = \max \| \mathcal{M}(D) - \mathcal{M}(D') \|_1$$

As seen in one of the examples in the preceding paragraph some mechanisms can have a very high or even unbounded sensitivity, which results in a very high amount of noise added to guarantee privacy. This in turn massively decreases accuracy. In order to improve the privacy-accuracy trade off different variations of sensitivity have also been defined, such as Smooth Sensitivity. Instead of accounting for all possible pairs of neighbouring datasets, Smooth Sensitivity is based on an analysis of the neighbouring datasets of the actual training set D, which reduces the amount of noise added [120].

Definition 6. (Smooth Sensitivity):
The smooth sensitivity of a randomized algorithm $\mathcal{M}$ for dataset D is:

$$S^*(\mathcal{M}, D) = \max_{k=0,1,\ldots,\|D\|} \left(e^{-\epsilon k} \max_{D' : \|D - D'\|_1 \leq k} \left(\max_{D'' : \|D' - D''\|_1 \leq 1} \| \mathcal{M}(D') - \mathcal{M}(D'') \|_1 \right) \right),$$

where ϵ is the privacy budget of $\mathcal{M}$.

Some prevalent and widely used mechanisms are the Laplace mechanism [57], and the Gaussian mechanism [18], for summary statistics. The exponential mechanism [114] is used when the outputs are discrete and it makes use of a utility function for responses. For optimisation based tasks, like regression analysis the Functional mechanism [179] is widely used.

Definition 7. (Laplace Mechanism):
Given any function f with co-domain $\mathcal{R}^m$, the Laplace Mechanism is defined as :

$$\mathcal{M}(D, f, \epsilon) = f(x) + (Y_1, ..., Y_m),$$

where Y_i are i.i.d. random variables drawn from $Lap(\Delta f/\epsilon)$.

The Gaussian mechanism is very similar but Y_i are drawn from the Gaussian distribution with $\mu = 0$ and $\sigma = \frac{\Delta f\sqrt{2log(1.25/\delta)}}{\epsilon}$ [18].

Definition 8. (Exponential Mechanism):
Using a scoring function $u(x, z) : u \to \mathbb{R}$, where u has a higher value for more preferable outputs (z). $\mathcal{M}$ ϵ-deferentially private if:

$$Pr(\mathcal{M}(D) = z) \propto exp\left(\frac{\epsilon u(x,z)}{2\Delta u}\right)$$

Returning to the mean height example discussed in the Introduction, the example box below showcases and demonstrates the technical terms defined in this section based on a Laplacian mechanism example.

Laplace Mechanism Example

Given D, our dataset made up of the heights of individuals A,B and C and the Laplacian mechanism then the mathematical notation for the Private Average heigh Query is:

$\mathcal{M}(D, f, \varepsilon) = f(D) + Lap(\frac{\Delta f}{\varepsilon}) = \overline{D} + Lap(\frac{\Delta f}{\varepsilon})$

The sensitivity for the general mean is:

$\Delta f = max\left(\| \frac{A+B+C+...+Z}{N+1} - \frac{A+B+C+...}{N} \|\right) = max\left(\| \frac{NZ-(A+B+C+...)}{N(N+1)} \|\right)$

Over all possible values of Z hence usually infinity, however as we are querying heights our domain is between 0 and 2.80 m and hence bound with maximum value 1.40 (with a dataset with one person with height 0 and the neighbouring dataset with an added person with height 2.80)

The maximum output difference considering our dataset D (a simpler "version" of smooth sensitivity), would be 0.4525 (the neighbouring dataset would have an extra individual with height 0), which is significantly smaller.

Given the Laplacian distribution expression we can infer that the smaller the sensitivity the smaller magnitude of the noise added and the smaller the privacy budget (which is the equivalent of more private) the larger the amount of noise added.

More complex differential private models and algorithms tend to be constructed by making use of DP properties and the base mechanisms described, like the algorithms discussed and implemented in Chapter 6. These complex algorithms are able to maintain good performance by efficiently diving the total privacy budget to each query and reducing their sensitivity. These complex DP models can range from classical logistic regression, to stochastic gradient descent (used in deep learning)[3], to GANs [62] as well as in the upcoming models of federated learning[89].

Apart from different models there are also quite a lot of variations, extensions and relaxations of DP, apart from Definition 1. Pejo et al [128]

summarises around 200 definitions and creates a system of knowledge to place the different definitions.

Thus far, we have considered DP generally, but there are in fact two models of differential privacy: the central model, the one we focus on in this thesis and that we refer to when using DP, and the local model which will be referred to as Local Differential Privacy (LDP). The difference between these two models is related to where the noise is added, i.e. in the central model the data from the different individuals is combined first and then the noise is added in one of the three separate stages discussed previously. In the LDP model, the noise is added by each individual before the data is collected and aggregated which means that the dataset holders never have access to the original data.

DP has many applications in various fields other than Privacy such as Robustness, Adaptive Data Analysis and Multi-Agent Systems [181].

2.2.3 Disparate Accuracy Loss

Due to the addition of noise to achieve privacy, there tends to be an associated decrease in accuracy. This is defined as the ratio of correct predictions (where Y and $\hat{Y}$ are the same) over the total amount of predictions in both this thesis and more broadly in the field. However, the accuracy decrease is not necessarily equal across subgroups, see simple example below in Figure 2.3 and Figure 2.4.

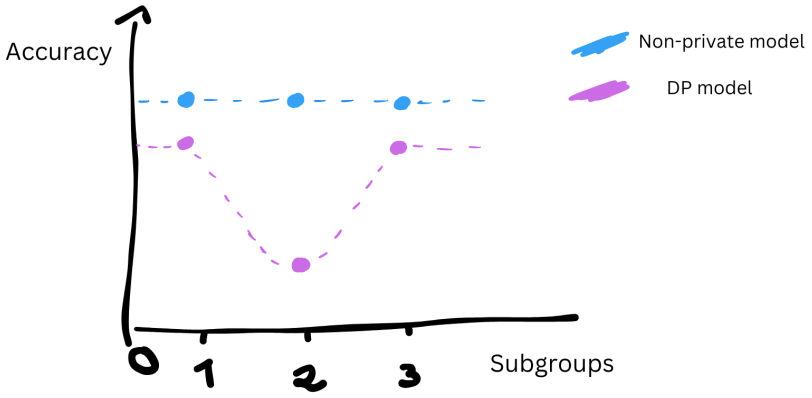


Figure 2.3: Sketch ofDisparate Accuracy Loss (DAL) performance plot example

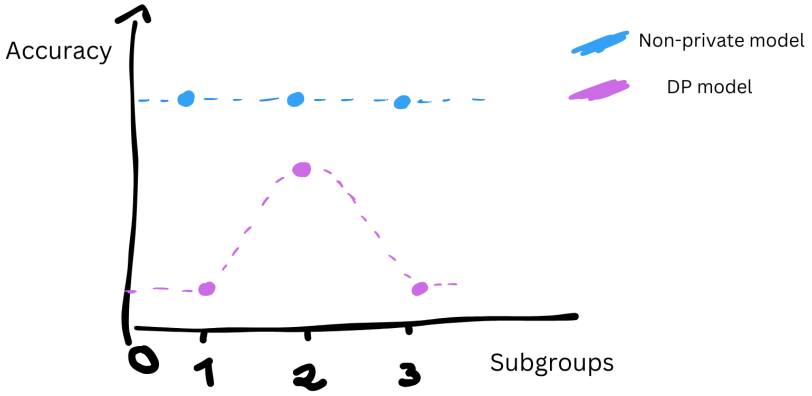


Figure 2.4: Sketch of Opposite Disparate Accuracy Loss (ODAL) performance plot example

In Figure 2.3, we represent a Disparate Accuracy Loss (DAL) for subgroup 2 compared with the rest of the dataset. Comparing the difference between the blue and the purple dots (subgroup accuracy for the non-private and private model correspondently) for each of the subgroups, we can see that the difference in group 2 is significantly bigger the rest of the subgroups. In Figure 2.4, the opposite occurs: the difference between the private and

non-private accuracy for subgroup 2 is significantly smaller than the rest of the subgroups. In the mathematical context of this thesis significantly smaller or bigger corresponds to a difference in order of magnitude, this is, ten times smaller or bigger. While the difference between these two figures in an abstract way does not appear to have much difference, when considering the context of loan allocations there is a big difference between DAL and ODAL, where in DAL there is a group that will be worse off by either not accessing credit products that they would in the non-private case or by being given access to product they cannot afford. In ODAL however there is a group where the decisions are more accurate, but for a differential private model to be implemented then the accuracy for the majority of the subgroups is at an acceptable level.

The algorithm which has the most amount of research on its accuracy drop behaviour is DP-SGD. Initial studies have shown conflicting results: Bagdasaryan et al.[14] found a bigger accuracy loss for underrepresented and complex groups; while Jaiswal et al. [95] arrived at a similar conclusion, they also showed that the biggest accuracy loss also occurred in the majority group. Xu et al. [176] explained that the disparate accuracy loss of DP-SGD is dependent on the gradient distributions, meaning, usually groups with a larger gradient incur a bigger drop in accuracy. The gradient distribution can be affected by the group sample sizes, but also other factors such as the model and privacy implementation used, as well as, the complexity of the data among others.

Farrand et al. [64] explores the impact of the data imbalance. The study shows that even small imbalances and loose privacy requirements can cause disparate impacts. DP-SGD has also been compared to different types of models, such as Private Aggregation of Teacher Ensembles (PATE) in Uniyal et al. [162]. PATE has been found (empirically) to create a smaller

disparate impact on the under-represented subgroups of datasets [162].

All the literature discussed thus far regarding the disparate impact of differentially private models has focused on classification and supervised or semi-supervised models. Ganev et al [75] focuses on the effects of differential privacy on generative models, i.e. models that learn the underlying probability distribution of the training dataset and then generate synthetic datasets. In terms of the effects on accuracy there is a disparate effect by all models tested (DP-SGD, PATE, DP-WGAN), however, in some settings DP models trained on the synthetic data generated perform better than on real data. All the effects described are exacerbated with stronger privacy guarantees.

The distribution of accuracy loss when implementing different differentially private models is still a highly understudied area of research, although it is currently gaining momentum. Chapter 6 aims to add to the literature focusing on a specific family of models, Decision Tree Based Models, which will be introduced in more detail in the next subsection.

2.2.4 Decision Tree Based Models

Decision Tree (DT) algorithms are:

- non-parametric (do not assume that the data follows an underlying normal distribution); and
- supervised (the dataset contains the true value of the quantity one is trying to predict, which is used in the training process)

This type of model can be applied in either classification or regression tasks this is, they split the dataset according to different values of covariates

until a classification or regression value is achieved [41]. DT are the most common type of algorithms used in advanced and critical development stages across the financial industry [13], hence why they feature as the focus of our technical study (Chapter 6). A simple example is shown in Figure 2.5.

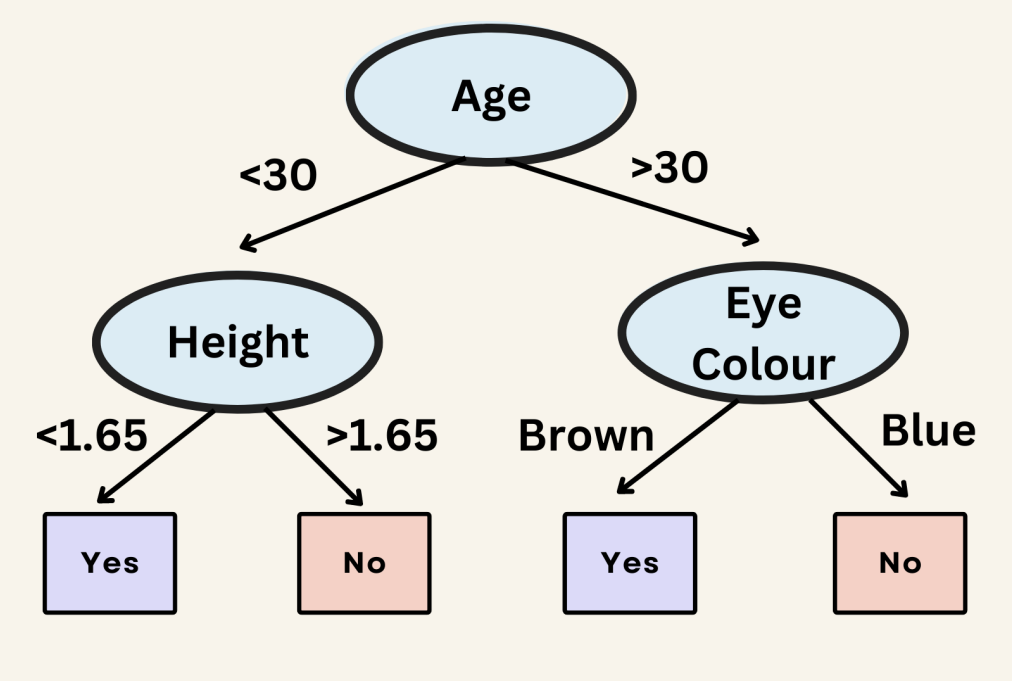


Figure 2.5: Example of simple Decision Tree. Lines are named branches, the ovals are nodes and the squares which the class are denominated leaf nodes. The number of layers in a tree is called depth.

Decisions trees have several advantages over different types of models as they are more easily interpretable (however this interpretability decreases for decision forests and other models such as GBMs), they have a non-parametric design, can handle discrete and continuous variables, have a low computational cost and can handle multi-class classifications and regressions [68]. There exist different types of algorithms based on decision trees, which mainly vary according to number of trees or the function used to split and create branches during the training process.

Algorithms with more than one Decision Tree are called Decision Forests,

these models are an ensemble of several individual DT which outcomes are aggregated according to the majority. A decision tree is formed by a recursive process of splitting of the data set into disjoint subgroups. For each split (which will lead to new child nodes or leaf nodes) a covariate is chosen and the following disjoint data sets will be based on the value of this covariate at each data point. For example, for a continuous variable, a splitting point will be chosen (how it is chosen will be described in more detail in the following paragraphs as it depends on the types of DT) and if a data point has a value below the splitting point it will go into one of the child nodes, if it is above it will go into a different child node, see Figure 2.5 for example of continuous and discrete covariates. This process stops when any of the termination conditions are met. Usually, these are related to the maximum tree depth, the nodes not having enough data to keep splitting into child notes or all covariates have been previously used in the tree [68, 66]. Some models can have some back-propagation that deletes leaf nodes which are not very reliable, this process is called pruning.

Models can either be Greedy or Random, depending on the rules for choice of attribute and splitting point when training. Greedy decision trees maximise an objective function when splitting nodes into new branches. Commonly used splitting functions are: Information gain [141], gain ratio [132] and Gini Index [29]. Random DT randomly choose an attribute at each node and a splitting point, the accuracy for single random trees is very low however Random Forests (models which aggregate different Random Trees) produce adequate results, the random approach to splitting significantly reduces the computational cost of the algorithm.

There are also other decision tree based algorithms like Gradient Boosting Machines (different way of aggregating the individual models) or Consistent Random Forest (split drawn from a probability function). Gradient Boost-

ing Machine (GBM) are also a combination of different simpler models however they are aggregated sequentially in an additive manner. The new models are trained based on maximising the negative gradient of the loss function of the previous model, using a gradient descent method. GBMs are very versatile models as the loss function can be chosen in order to best fit the task at hand, are relatively easy to implement have produced successful models in practice [118].

Differentially Private Decision Tree Based Models

Due to the characteristics and properties of DP, not all efficient non private DT based models will be efficient when converted to DP. Fletcher et al [68] looks into how differential privacy might affect the different components of this family of models and review the existing models in the literature.

The determinant factor of how well an algorithm performs when adapted for differential privacy is related to how well the privacy budget is distributed across the different queries and computations that involve access to the training data. The less queries of the data that are performed, the less the total budget gets used and hence the bigger budget per query we can allocate, which in turn leads to a smaller addition of noise and hence potential better accuracy.

There are different types of data queries that happen during training of a DT based model. Non-leaf node queries are the queries related to the optimisation of the splitting function and hence are only present in Greedy trees. Leaf queries tells us which is the majority class are present in all DT based models. In order to add less noise it is possible to reduce the sensitivity of both these types of queries by adopting local or smooth sensitivity however this comes at the cost of weakening the privacy definition. De-

pending on the termination criteria used it may or may not involve a data query, which needs to be taken into consideration in the privacy budget allocation. Different approaches have been used in terms of the trade-off of precision of termination criteria versus the amount of noise added due to data queries. Finally, pruning can also require data queries, however, some algorithms make use of past query results for this, avoiding further exhausting the privacy budget.

For more theoretical work on the effect of Differential Privacy on DT based models and experimental algorithm comparison see [68]. In the Algorithms and Models Subsection of Chapter 6 the models chosen to be evaluated will be examined in more detail.

2.2.5 Usable DP and Communicating DP

Apart from academic work, DP has started to be implemented in different industries and governments. LDP has been implemented by Google to analyse browser settings data, by Apple to collect emoji and word usage data on iOS 10 and macOS 10.2, and by Microsoft [175]. Uber has also implemented DP to prevent data analysts from stalking customers and the US Census Bureau is using DP to prevent information disclosure in the 2020 Census [46], furthermore, DP is also used by Meta and LinkedIn (owned by Microsoft) [98].

As a result of these implementations, research on the communication of DP and usable DP (a concept based on usable privacy notices, I.e. the information about DP should be understandable to a layperson so that they can make informed decisions) has grown exponentially over the last couple of years.

The studies within this area can be separated into two groups based on the method of explanations: textual communication methods and visual communication methods.

Starting with the textual explanations, Xiong et al [175] and [103] tested different textual explanations, which were designed based on the communications of institutions that have implemented DP. They both found that participants are more willing to share their information with the implementation of DP (where data sharing was larger with DP over LDP), however that the descriptions of DP were hard to understand, specifically concepts around randomness and noise. Kuhtreiber et al also found that the differences between DP and LDP were not well understood and the authors suggested the trial of more visual communication methods in the future. Cummings et al [46] expand on this work by also taking in consideration users' privacy expectations in their willingness to share data. This work is one of the first that starts to consider the voice of the user as it explores how DP satisfies their expectations. This work found that users care about the types of information leaks DP can protect against and that they are more willing to share their data when the likelihood of a leak is lower.

Regarding visual communication methods, Bullek et al [33] first studied how users understand Randomised Responses (a mechanism commonly used in LDP) by making use of a visual representation of a different colour spinner to better represent the random aspect and correspond probability in a more understandable way. Karegar et al [98] created different visual metaphors for both central and LDP in an iterative design process involving feedback from privacy experts. The metaphors created also examined the privacy budget and the privacy-accuracy trade-off. As a result of this research activity, they further formulated a functionality list to evaluate DP communications. Wen et al [167] found that visual explanations of LDP

improve participants' understanding better than textual explanations, as previously speculated by Kuhtreiber et al [103]. The authors further speculate that the improvement is due the visual analogy of the mathematical concept of privacy to a lottery draw.

From all the studies defined above apart from Karegaret al [98], none of the explanations account for the effect of the privacy parameter (equivalent to privacy budget). Smart et al [150] discuss the importance of including the privacy parameter in DP communications as failing to address this can lead to a false sense of "security" for user, I.e., thinking their data is more protected than it is, and hence hindering users' capacities to make informed decisions, and could lead to a legitimisation of broader data collection, a concern also raised by Sarathy [144]. Sarathy also expands on the impact of the formalisation of privacy by using mathematical concepts:

> "Overall, I find that a privacy discourse dominated by any single – and mathematically powerful – standard risks ascribing the role of privacy protection solely to technological artifacts, rather than to the social, political, and economic orders that are co-produced along with these technologies. Indeed, this paper argues that reducing the broad and multi-faceted nature of privacy to a narrow yet alluringly elegant technical definition is part of the motivation for Big Tech companies to adopt differential privacy; it allows these institutions to achieve closure of the privacy problem without changing their underlying values and practices." [144]

These two points are each partially addressed by Nanayakkare et al [117] and Aslan et al [8]. Nanayakkare et al developed three different methods to explain how the privacy parameter works. The study found that odds-based

explanation methods are more effective than output-based methods [117]. Aslan et al [8] implemented two DP algorithms for healthcare datasets, in order to evaluate their privacy-accuracy performance in an applied context.

Aslan et al's study is one of the first in the literature to highlight the need to study the sociotechnical impacts of DP within applied contexts, similar to the work in this thesis, however, it is mainly focused on the technological aspects, with the sociotechnical aspects being left for future work. User input and attitudes towards the privacy-accuracy trade-off and different accuracy behaviours is still not addressed in the literature, however, the acknowledgement by Aslan et al. [8] and the literature on usable DP, indicate a turning point, of which the work of this thesis on the sociotechnical impacts of DP implementation in the Credit Industry is a part of.

By having studies that focus on user and industry attitudes towards DP, we can start unpacking the political and social implications of DP, i.e. do users think DP addresses their personal privacy concerns, how is DP related to the institutional logic of the credit industry and how does this impact its potential implementation.

2.3 Perceptions, Sensemaking and Attitudes to Algorithms

Consumers' experiences with loan applications and the consumer credit industry, cannot be dissociated from the embedded ML technology which underpins all application processes and decision-making. As algorithms are becoming ever more ubiquitous in our daily lives, so is the need to under-

stand how people perceive, make sense and interact with them and the outputs derived from them, e.g., the decisions of a loan company regarding which applications are successful. The areas of algorithmic sensemaking and FACcT (Fairness, Accountability and Transparency of algorithms) perception research have started addressing these questions.

Sensemaking refers to the way people ascribe and derive meaning to or make sense of their experiences [54]. It is an interpretative process which depends on pre-existing understanding and attitudes which are updated as a result of sensemaking of the experience. The concept of a mental state which influences and is updated with experience also corresponds with part of Vickers' concept of Appreciative System [111, 36]. Applied to algorithms, sensemaking refers to the way people interpret decisions made by the algorithm. It is through sensemaking that consumers create a mental model, which in Human-Computer Interaction (HCI) literature refers to what consumers believe about a system and consequently what they base their perceptions of technology on. A mental model affects how they interact and use the technology, hence well-designed technology should provoke mental models which are close to the intended workings of the technology [121]. It is important to understand how users make sense of algorithmic tools in order to design a better user experience which is characterised in part by a small gap between MM and process. Design choices impact end-users' sensemaking and acquired knowledge [82], a factor which has the potential to positively or negatively affect user experiences of loan applications.

In the context of consumer credit and loan applications the mental models of the process is not just referring to a consumer's interactions with a lender's platform (which is essential for an intuitive application process) but also an MM of the decision-making behind the outcome. Both of which shape how the user interacts with the financial sector.

Just and Latzer [97] argue that algorithmic applications form and construct realities, similarly to mass media, and that users' sensemaking and interactions with algorithms is an interdependent and cyclic process. Their findings are supported by Shin and Park [148], whose study focuses on users' perceptions of FACcT and its relations with Trust and Satisfaction relating to news media recommendation systems, finding that perceived FACcT plays a significant role in user satisfaction and that trust plays a moderating role in the effects of FACcT on satisfaction with service.

Overall, the literature finds that perceived algorithmic fairness tends to be subjective and contextual [148, 17, 106]. Furthermore, Baleis et al. [17] find by process of a systemic literature review that algorithms tend to be seen as fairer when referring to mechanical tasks e.g. processing quantitative data for objective measures. Schöffer, Machowski and Küh [147] found mixed attitudes regarding automation where those in favour saw it as more objective, those against found it lacking empathy and some participants acknowledged this trade-off.

Notions of fairness and bias are a particular concern in the context of loan applications. Saxena et al. [146] explores which definitions of fairness the general public prefers (out of three fairness options based on distributive justice) via an experimental design. The study concludes that there is some support for affirmative action, a set of actions to improve opportunities for members of groups that have been historically discriminated, e.g. race and other sensitive attributes. "The Perception of Fairness of Algorithms and Proxy Information" report by the Behavioural Insights Team [155] found that people have a negative perception of algorithmic decision-making when compared to other methods. However, when people thought that the algorithm was more accurate, it was perceived as fairer than an algorithm that was described as less accurate. The study also found that the use of proxy

information in this scenario (for gender, race and social groups) was seen as unfair.

Wang, Harper and Zhu [166] found that perceptions of fairness strongly increase with a favourable outcome to the individual and with the absence of bias at a group level (where the first effect is bigger than the latter). The impact of favourability (or unfavourability) outcomes on fairness perception diminishes with additional years of education.

In summary, whilst the use of ML technologies is increasing within the UK financial sector, user understandings of ML decisions as well as their attitudes towards the fairness of ML tools are variable. In the interests of considering how to better design this process in a way that avoids propagating perceived unfairness, it is important to explore user experiences of automated loan application services in use today.

2.4 Summary

This chapter briefly expanded on three different areas of literature, which are brought together in the studies presented in the following chapters.

The first area is focused on the UK financial industry, specifically the consumer credit industry. It starts by highlighting the role of governmental policies and evolution of the technology in the industry. It then explores the technology used today based on a series of governmental reports, this section of the literature is mainly based on quantitative survey data and hence does not delve onto the reasoning behind the choices in this technology, which is addressed in the first part of the Industry consultation study in Chapter 5.

The second area of literature was focused on DP, starting with its technical element, which will be used in Chapter 6, in the implementation and evaluation of different differentially private decision based tree models. The novel area of Usable DP is then quickly summarised, it is within this area that my work finds its home. My work expands on the call for a sociotechnical and consumer-centred approach to the study of DP. The findings from this area also massively influenced the design of the interactive board game activity presented in Chapter 7.

Finally, the last section draws from the FACct and HCI literature and focuses specifically on users/consumers experiences, sensemaking and attitudes to algorithms. This literature serves as the basis for the design of the initial exploratory interview study towards consumers experiences of loan applications in the UK, presented in Chapter 4.

In the next chapter I will discuss the relevant methodological literature and the methodological choices of this thesis.

Chapter 3

Methodology

The initial section (3.1) of this chapter describes the lens and positioning of this thesis, such that the studies presented, and their results can be interpreted within this context. The next section (3.2) explores different methodologies generally in order to understand the choices of methods chosen and their implications. The third section (3.3) focuses on the specific qualitative methodologies used. Finally, the last section (3.4) is the researcher's reflexive statement.

This chapter builds on the thesis overview of Chapter 1, discussing methodologies and approaches used and why. The next chapters will focus on the research activities themselves and the findings' interpretation and discussion in light of the methodological framework described here.

3.1 Ontology and Epistemology in Research and Technology

Different researchers and academic fields have different underlying worldviews [45], these beliefs shape the way individuals perceive and act in the world [83]. A researcher's worldview encompasses their assumptions on what exists in the world that we can acquire knowledge about (ontology) and how one creates knowledge (epistemology) [135]. The ontology and epistemology of a researcher can be generally positioned within a worldview/philosophical orientation spectrum [135]. A researcher's worldview and position on the philosophical orientation spectrum can be shaped by their discipline, past research and life experiences, being able to change and evolve over time [45].

On one side of the spectrum, we have positivistic views, which see reality as being pre-existing and external to the subject (realist ontology). In terms of epistemology, this view sees knowledge as universal, objective and measurable. A researcher with this worldview acts as a disinterested observer [135].

Positivism is based on an objective external reality which can be directly observed and measured [38]. It is based on finding causal relationships, and research usually reduces general ideas to a smaller set of hypotheses with associated variables and research questions which can be tested using observation [45]. As such, meanings and experiences are outside of the scope of knowledgeable truth [38, 60, 81]. Furthermore, knowledge such as, for e.g. the existence of black matter and other purely theoretical findings are not seen as valid truth as these are unobservable. The incompatability of theoretical findings led to the development of a worldview named

post-positivism which tends to underpin most current empirical research [38]. This worldview acknowledges validity of truth based on unobservable entities due to their effect on observable and measurable entities.

In positivism researchers are seen as having the ability to be completely independent from the research, hence being seen as neutral observers, while under post-positivism there is an acknowledgement of researcher and theoretical biases and hence findings are not interpreted as absolute truth but under a probability of the knowledge holding in similar cases, therefore research has an aim of approximating findings to the truth [38].

On the other end of the philosophical orientation spectrum we have constructivism where reality is socially constructed by each individual, hence internal and multiple (relativist ontology) [135]. Based on this view of reality knowledge is then subjective, particular and contextual which leads the role of the researcher to be that of a participant interpreter.

The constructivist worldview is based on local and specific realities where individuals seek an understanding of the world through meaning-making activities of groups and individuals [45, 108] . The researcher, as a participant interpreter seeks complex views and ideas rather than narrowing meanings to a small number of concepts [45]. The research realm tends to be about human experience and social phenomena.

While researchers on the end point of the philosophical position spectrum see these positions as irreconcilable and hence their standard associated methodologies as well, i.e. quantitative and qualitative methods, such as Lincoln and Guba [108], there are others that argue against this. Willig [169] states that because of collapsing ontology and epistemology together, constructivism is inevitably associated with relativism, however, most constructivist qualitative research has realism as an underlying ontology ar-

guing that ontological relativism is not compatible with research inquiries. Cupchik [47] also highlights that the internal and external divisions under which these ontological divisions arise from are not as straightforward as assumed when discussing the topic. Cupchik also states that while positivistic and constructivist researchers approach the phenomenon differently, they can agree on the existence of social phenomena independently of the existence of researchers and that both approaches can complement each other:

> *"This interplay between descriptive richness and experimental precision can bring accounts of social phenomena to progressively greater levels of clarity. Together, qualitative and quantitative methods provide complementary views of the phenomena and efforts at achieving their reconciliation can elucidate processes underlying them. Constructivist realism is an ontological position that accommodates the best of positivism and interpretivism."[47]*

It is within a constructivist realist approach that the research in this thesis is situated, as it best reflects my philosophical positioning. In the next section of research methods, I will discuss how this shapes the methodologies used.

Apart from the worldviews already mentioned there are others which shape research inquiries, such as pragmatism (not committed to any specific epistemological and ontological stances) and the advocacy and participatory worldview (research should address issues of social justice) [45].

As previously mentioned, certain disciplines tend to lean to one of the sides of the philosophical spectrum. In the case of finance and economics, most

work is done from a more postpositivist approach to research, basing it on mathematical modelling and either mechanical or biological principles [105, 48]. Throughout time there have been authors such as Marx and Keynes that have taken different approaches, but since the financial crisis of 2008 there has been an increased attention to the limitations of these mathematical modelling approaches. Lawson [105] argues that these mathematical modelling tools are not adequate to study the social phenomena aspect of economics and defends a wider inclusion in the academic field of other non-formal approaches to economics and finance.

The financial literature of the previous chapter starts addressing some of these points. Based on the field of the sociology of finance, the work discussed in Chapter 2 and the basis for the research studies in the following chapters takes a non-formal and more social constructivist approach to the study of financial institutions and concepts.

People's underlying ontological and epistemological views also influence their understanding and views of technology and its place in the wider context of society. Positivistic views of technology are associated with technological determinism which is the view that technological changes force social adaptations and hence influence history [149]. Since its inception, the idea of technological determinism has been criticised by science historians and sociologists as an over-simplification of historical changes based on one factor (technology) while overlooking others, such as social and cultural factors [55].

From the more constructivist side, the theory of the social construct of technology (SCOT) is one of the big models. This theory states that all knowledge develops as a result of social interaction and language use, hence being shared experiences, to directly oppose Technological Determinism. The

SCOT model rests upon interpretative flexibility, which in the scientific context refers to different scientists' interpretations of the same results, for example, to construct scientific knowledge. In the technological context this refers to the way the technology is interpreted and constructed in different ways by different social groups. SCOT further denies the linearity of the innovation model, which tends to be described with historical retrospective, focusing mainly on the specific variation of the technology, which was deemed successful [129, 164]. One of the criticisms of SCOT is its lack of acknowledgement of the materiality of technology [149, 145].

There are a series of different perspectives and theories which acknowledge both the social and material roles within and of technology, one of which being the sociotechnical premise. The sociotechnical premise is defined by three characteristics:

- the mutual constitution of people and technology
- the contextual embeddedness of the mutual constitution
- the importance of collective action

The mutual constitution refers to the intertwined and indivisible relation of the social and the technological without making judgements on the relative importance of each of the components, as ultimately these are indissociable. The second requirement refers to the dependence of the sociotechnical on its context, where context is also seen through a dynamic and holistic lens. Finally, the third defining characteristic stresses the importance of collective action, which in this context refers to the pursuit of a shared goal by several stakeholders, where at times different stakeholders can pursue different goals creating conflict. This collective action shapes the design,

development and implementation of different technologies, and leads to a complex view of social settings. This can be summarised in:

> *"The premise of collective action is that joint interests and multiple goals are intertwined with both the context and the technological elements" [145].*

In my work, I take an explicit sociotechnical approach, which matches my constructivist realist position as I see the world from a holistic and interconnected perspective acknowledging both the constraints of the physical materiality and the social construction of knowledge and experiences.

Figure 3.1 summarises the two main positions on the philosophical orientation spectrum, placing the main technological theories discussed in the chapter as well as the researcher's philosophical orientation.

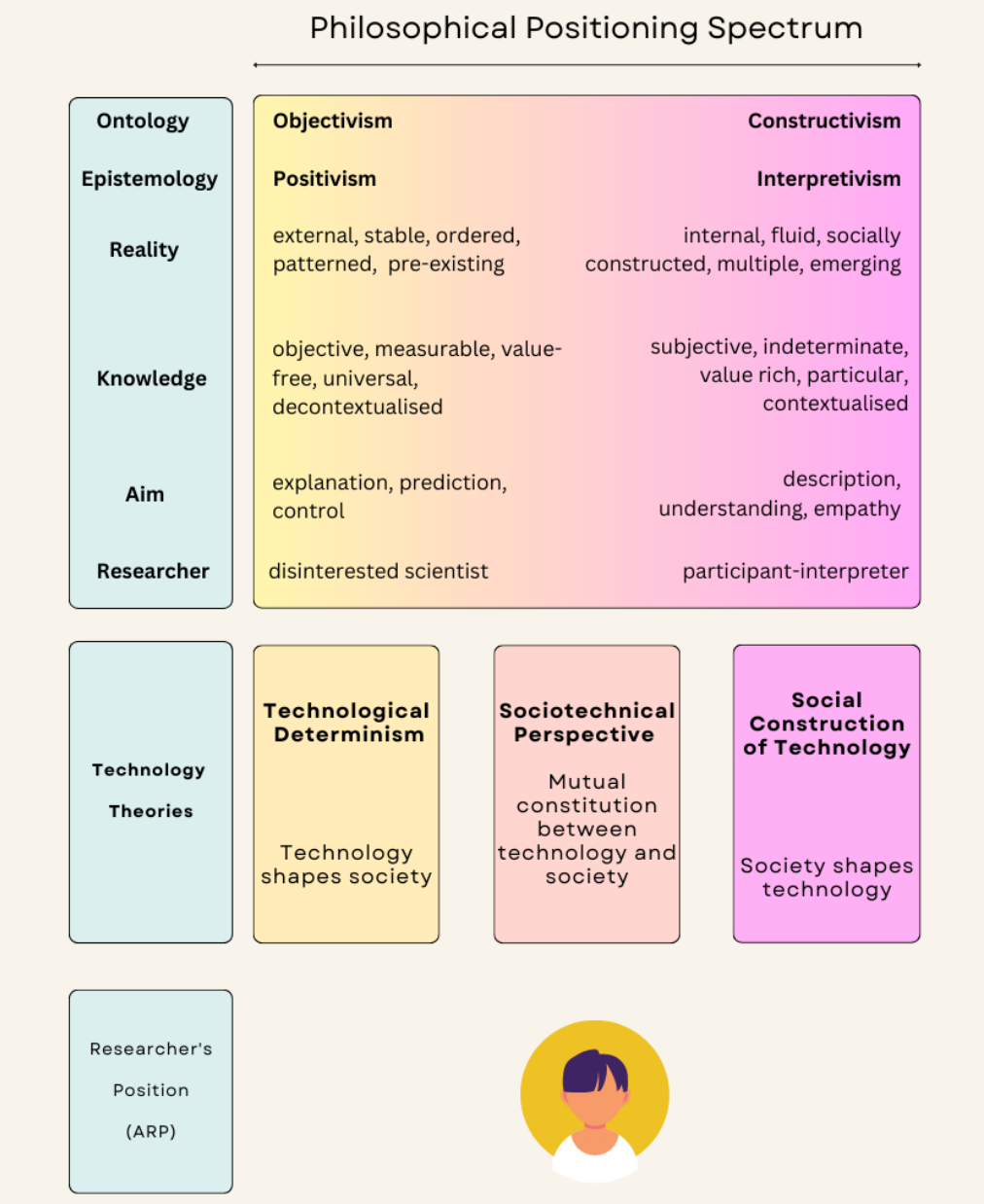


Figure 3.1: Philosophical Position Spectrum based on Raqib et al. [135]

This thesis is written in the first person to keep with the philosophical position of the researcher as a participant-interpreter. Further, it is done to acknowledge the role of the researcher in shaping the findings and pursuing reflexivity (the explicit examination of one's own beliefs, judgments and practices during the research process and how these may have influenced the research) within the research work, see more in section 2.4.

3.2 Research Methods

In this section I will review the variety of methods used in research studies, from a theoretical perspective. The choice of methods is related to the research question (which in turn reflects the researcher and academic discipline epistemology). For example, natural sciences tend to employ experimental and quantitative methods; social sciences tend to use a combination of qualitative and quantitative methods. In the next section, Study Methods, I will use the content of this section to describe and justify the choice of methods for the research activities present in this thesis.

When research questions are related to experience, meaning and perspective, from the standpoint of the participant, qualitative methods are the most adequate [85]. These data are usually not amenable to counting or measuring, which would be the case for data collected from quantitative methodologies, i.e. surveys. Quantitative methods are employed to predetermined hypothesis, while qualitative methods are often associated with a more exploratory approach to research[25]. While qualitative and quantitative approaches typically address different types of questions, they often complement each other in "mixed methods" studies.

Usually, qualitative methods tend to be associated with more constructivist worldviews and quantitative methods with more positivist worldviews, both these family of methods as well as mixed-methods approach can also be based in different philosophical perspectives.

3.2.1 Qualitative Data Collection Methods

Within the broad family of data collection qualitative methodologies there are a variety of methods: observational studies, interviews and focus groups, document and discourse analysis, among many others.

Some of the most commonly employed qualitative data collection methodologies are interviews and focus groups, these two methods will comprise the three qualitative research activities in this thesis. Interview-based methods allow researchers to explore individual participants' attitudes, views and experiences. Interviews are also a good method for the exploration of past events through retrospective questions and provide rich datasets [31, 45].

Interviews can be of different types from structured where the interviewer follows a very strict set of questions in a specific order to semi-structured where the interviewer has a set of key questions but also pursues topics that might arise from the conversation with the participants to unstructured interviews which do not have pre-planned questions [27, 45].

To conduct effective interviews and be able to collect rich data, the interview design is essential. When creating an interview guide the researcher should consider the choice of language, the necessary questions to answer the research question, possible follow-up questions and probes as well as the ordering of the questions [154, 53]. Choosing accessible and simple language, and starting the interview with easier and more general questions helps establish a report between the interviewer and participant [53, 26]. All these elements should then go through some pre-testing to identify potential improvements to the interview guide.

When running the interview study the moderator should always start by

outlining the purpose of the study and reassuring participants that there are no correct answers and remind them of their rights to skip or stop the interviews at any time, this being particularly important for interviews covering sensitive topics such as personal finances [49]. A good interviewer should be a empathic listener and be able to know when to ask probe and follow-up questions when necessary give the interviewee space and time to reflect on the questions asked. It is also essential for a good interviewer to be able to get a good balance between letting participants bring up the topics that come to mind and when to lead the conversation back to the research topic [154]. These are skills that one develops with time, experience and critical reflection on interview transcripts.

Some of the short-comings of this family of methods are associated with the data collected being from the perspective of the participant. In other words, participants might be inclined to shape their answers based on their perceptions of what the interviewer wants to hear (confirmation bias), participants choose to respond and engage with the recruitment materials might have more charged views regarding the interview topic (selection bias and extreme response bias)[30].

Focus groups are similar to interviews but differ in the number of participants. This type of research allows participants to discuss amongst themselves and gather group opinions on given topics advanced by the interviewer/moderator. As a result of group dynamics group opinions can be categorised as divergent, when individuals interacting in a group do not always collapse to a single viewpoint, polarized, when average group opinions can become more extreme after discussions than they were to begin with and finally consensus where individual participants' views converge in the course of the focus group [116]. Compared to interviews focus groups are often a more time efficient way to gather different participants views

but at the cost of decreased depth with each participant. However, due to the nature of group processes, the data collected and expressed by participants tends to have more consensus view expressed in focus groups. This is a phenomena well documented in the literature [137, 152]. Furthermore, participants might feel less comfortable disclosing personal sensitive information in a focus group setting when compared to a one-on-one interview.

For both interviews and focus groups, a variety of stimulus materials and activities can be integrated as part of the research inquiries to aid in achieving the research objectives. A commonly used approach in contemporary social science is vignettes. Vignettes usually present a fictionalised event which is shared with the participants to incite participants to discuss their thoughts openly on the situation and gather their attitudes towards the topic [142]. Within the design field scenarios and personas are employed. A persona is a fictitious user based on data and a scenario comprises the story about the persona using a system that has not yet been created, similar to vignettes. Using personas, vignettes and scenarios provides a tool to see the situation from the user's/persona´s perspective and also develops a shared understanding of a concept from which discussions then occur, especially in applications where there is variation in experience or prior knowledge. [119, 158].

The use of vignettes can be particularly beneficial when asking participants to discuss sensitive topics, this is, they allow participants to express their thoughts on the topic without having to disclose personal information/and or involvement [142, 158]. If focused on a subgroup of the population over specific events, personas are also equally used in this context [119].

Other stimulus material shown to participants can be of all media types depending on the relevance to the research topic [158]. A less common

tool is games, but these have been used in Edwards et al. [61] to transmit information and gather attitudes towards complex systems.

There is a significant body of research on games, mostly as the research objects. Serious games, a term established in the 70s are a subset of games that have as a primary aim educational purposes instead of an amusement purpose[104]. These have had a range of application areas from military training to wellbeing, healthcare, education and cultural heritage among others [104]. Serious games, specifically role-playing games, have also been used as tools to study governance in complex systems [61] due to their capacity to transfer large amounts of technical, context and processed-based knowledge, as well as helping players understand the implications of said knowledge [140].

An interactive game board was designed to communicate DP with consumers and gather their attitudes towards the technology in the risk assessment model of the loan application, discussed on more detail in Chapter 7.

Another qualitative data collection methodology is ethnography. Although not specifically used in this thesis, ethnographic methods usually involve long-term observation in the field and can be hard to gain access to subjects but provide rich datasets. [9] Due to the nature of the methodology traditional ethnographic work is usually constrained by physical location. More recently ethnography has also been applied in an online context, with the creation of digital ethnography which focuses on online communities and also provides avenues to study harder-to-reach groups [99]. Observational methods are more adequate if the researcher would rather have first-hand experience of the participant and focused on specific discrete timeframes, over participants' views and perception (like in interview studies).

Within the context of Finance, there has been an interest in ethnographic methods however traditional ethnographic work is challenging due to the "opaque, secretive and increasingly placeless" nature of the industry [157] and where digital ethnographic methods also face their difficulties as the technologies and communication methods used are not public and integrated within the physical world. Tischer [157] proposed the use of open access financial documents as alternative points of access to organisations.

Documents such as institutional documents and communications, databases, websites etc. are often used as sources of data, in processes such as content analysis and are sometimes labelled as secondary data [151, 168]. This type of data is not strictly textual but can also be of a visual/photographic nature, video or audio [151, 168]. Using documents and artifacts as data sources bypasses the issue of physical constraint location such as in ethnography but also has some difficulties in gaining access dependent on the context of the research (especially when relating to private organisations internal documents).

3.2.2 Technical Research

Most machine learning research such as the one presented in section 2.2. does not follow either of these approaches. Within the field of computer science there is a very wide variety of approaches to research and associated philosophical positions. There are three main intellectual traditions: the theoretical, the empirical and the engineering based one [156]. The theoretical approach is focused on creating hypothesis and theorems and proving them, the empirical in forming models for predictions, experimenting and collecting data and finally the engineering approach in designing, implementing, and testing different systems to solve problems [156]. Often

these different approaches are all used within the same research area, for example in Chapter 2.2.3 on the literature regarding DP's disparate accuracy loss literature: Bagdasarayn et al. [15] and Jaiswal et al [95] are purely empirical studies, differing on the datasets implemented and with different results. Xu et al [176]then builds on these last two studies and using a theoretical approach provides an explanation for the differing results.

The confluence of different perspectives on the same research topic, which often refer to each other, combined with the lack of disclosure of research methods used [163] makes it hard to evaluate the validity and assumptions under which the findings hold. Computer Science (CS) research tends to take a de-contextual approach, this is research is not focused on specific contexts and their specificity as most positivist and post-positivist approaches to research. Nonetheless CS findings tend to be extrapolated to applied contexts. In this thesis my work brings a qualitative perspective to CS research, focusing on situated knowledge and the mutual constitutions of technology and societal institutions.

3.2.3 Mixed Methods and Triangulation

A variety of research projects and inquiries takes a mixed-methods approach, this is, it combines methods both from the quantitative and qualitative traditions [30, 45, 38]. A mixed-methods approach indicates an acceptance of the diversity of truths reachable through different methods [38]. Combining different data collection approaches via triangulation can help address each of the individual methods' limitations [45]. For example, combining interviews and surveys can address the lack of generality and small sample number regarding interviews and lack of depth regarding surveys.

Apart from methodological triangulation described above, there are also other types of triangulation: researcher triangulation, theory triangulation and data source triangulation [160]. Generally, triangulation refers to the use of multiple methods and/or data sources to develop a comprehensive understanding of the research topic in cause [127], it is seen as strategy to test and validate the convergence of information from different sources.

This thesis will make use of triangulation at different scales, from methodological in Chapter 4 - combining online interviews and surveys- to data source triangulation based on different studies described in Chapter 8 - where the findings from chapter 4,5,6 and 7 will be combined to answer the research question. The studies' Methods section of this chapter as well as the thesis structure already discussed in Chapter 1 highlight the benefit and need of combining different methods involving different stakeholders. This provides an optimised approach from which to investigate the complex system that is consumer credit and understand the possible consumer impact of the implementation of differential privacy in the risk assessment model of a loan application and thus answer the research question.

3.2.4 Data Analysis

There are a variety of methods for qualitative data analysis, where the choice of method depends on the nature of the data itself as well as the philosophical positioning and academic field of the researcher/research. Some of these methods include: content analysis, thematic analysis, narrative analysis, grounded theory analysis and discourse analysis. Content and thematic analysis are to some degree similar where both focus on analysing the meaning of the dataset, where content analysis is a systematic, rule-guided techniques to analyse the informal contents of data [113] and the-

matic analysis finds patterns and generates themes [27]. Narrative analysis also focuses on the meaning of the data but from the perspective of individual participants narratives and analysing differences in cases [178] using structural devices such as plot, setting and activities [45]. Grounded theory analysis generates new theory based on the data analysis. The process of data collection, data analysis, and theory development happen in an iterative process. Iterative data collection and analysis occurs until one reaches theoretical saturation, the point at which additional data adds no additional insight into the new theory, therefore this method is best used when there are no existing theories in the literature [43]. Finally discourse analysis focuses its attention not just on the content of the text but on the textual choices themselves, this is, most forms of discourse analysis aim to provide a better understanding of socio-cultural aspects of texts, via socially situated accounts of texts, being associated with constructivist views and focused on linguistics [102].

In the rest of this section I will discuss the family of Thematic Analysis in more detail as it is the method that I use in the three qualitative studies presented in this thesis (Chapter 4,5 and 7). I will also briefly outline my personal analysis method and its evolution over the course of the PhD.

Thematic Analysis is a family of related methods under the following general subgroups: coding reliability, codebook, reflexive TA, and thematic coding [28]. Each of these methods tends to have different underlying philosophical positions. Coding reliability TA focuses on procedures for ensuring the objectivity, reliability or accuracy of coding, hence having a positivist perspective. On the other side of the spectrum, researcher that take a constructivist perspective tend to use reflexive TA where the researcher's subjectivity is a resource instead of being seen as a bias to be eliminated, this genre of TA rejects the notion that coding can ever be

accurate. Codebook TA (the group under which the Framework Method falls) is situated somewhere in between those two positions having some structured coding procedures as well as reflexive components. Finally, thematic coding is associated with a specific theoretical position, grounded theory [28].

The Framework Method has been used since the 1980s in the social policy realm but is ever more used in multidisciplinary contexts [73]. This method is based on two concepts: the analytical framework (a set of codes (in our context groups) organised into categories that have been developed by researchers involved in the analysis that can be used to manage and organise the data) and the framework matrix (a spreadsheet contains numerous cells into which summarized data are entered by codes, in our context groups, (columns) and cases (rows), in our context participants). Gale et al [73] describe this process, which is based on seven steps, in a research team context. The steps are as follows: transcription, familiarisation with interviews, coding, developing a working analytical framework, applying the analytical framework, charting the data into the framework matrix and finally interpreting the data.

Three of the four research activities in the PhD are of a qualitative nature and have been analysed via thematic analysis (TA) drawing from the Framework Method with an experiential orientation as it aims to capture participants' experiences and perspectives and ground research in participants' accounts.

My approach to TA has evolved and changed during the PhD, as I grew more accustomed to qualitative methods both because of PhD work and personal involvement in other qualitative research projects. Chapter 4, my first qualitative study followed a more traditional 6 step Braun and Clarke

TA done in Nvivo. This first analysis lacked some depth and nuance due to my inexperience- the analysis just followed the steps without experiential knowledge of what they meant. This analysis was then revisited at a different time point (few months later) when I had gained some more maturity as a qualitative researcher. Chapter 5 and 7 were then analysed following a slightly different approach for TA. I had learned through experience that my new approach worked better for me, as a tool to be able to look at the data in depth and create themes more easily.

My final personal approach to thematic analysis, draws from the framework method as well as general steps of thematic analysis [27], differing due to the analysis primarily being done by one researcher (ARP) and being discussed with other researchers (AL and MH) who are experts in qualitative analysis. The steps used were:

Transcription: Checking the automated transcriptions with the audio files manually, further served as familiarisation with interviews

Coding: Each text segment of interviews was coded and indexed (participants and location on transcript). The indexing of these codes means that step 5 of applying the analytical framework of the Framework Method is not necessary as it is combined with developing an analytical framework

Grouping/Developing an analytical framework: aggregating similar codes and giving them a group name (where the group name is what Gale et al [73] call a code), as well as reviewing the groups/codes developed based on the interview data.

Charting data into the framework matrix: By aggregating groups (with codes and indexes) into initial themes (or categories in Gale et al [73]) and creating a spreadsheet for each. The resulting document becomes

the equivalent (informationally wise) to a framework matrix and is then used to clearly visualise and read the large amount of data. At this stage, the framework matrix is shared with the expert researchers for discussion on the initial themes.

Interpreting the data/Defining and naming themes: After the initial themes have been developed, the researcher goes through the framework and corresponding interview transcript sections (via the indexing system) to further establish patterns and interpret the data in an interactive manner (where steps 3-5 can be repeated when and if necessary).

Figure 3.2 shows graphically the differences and evolution of my TA process from the three qualitative studies present in this thesis.

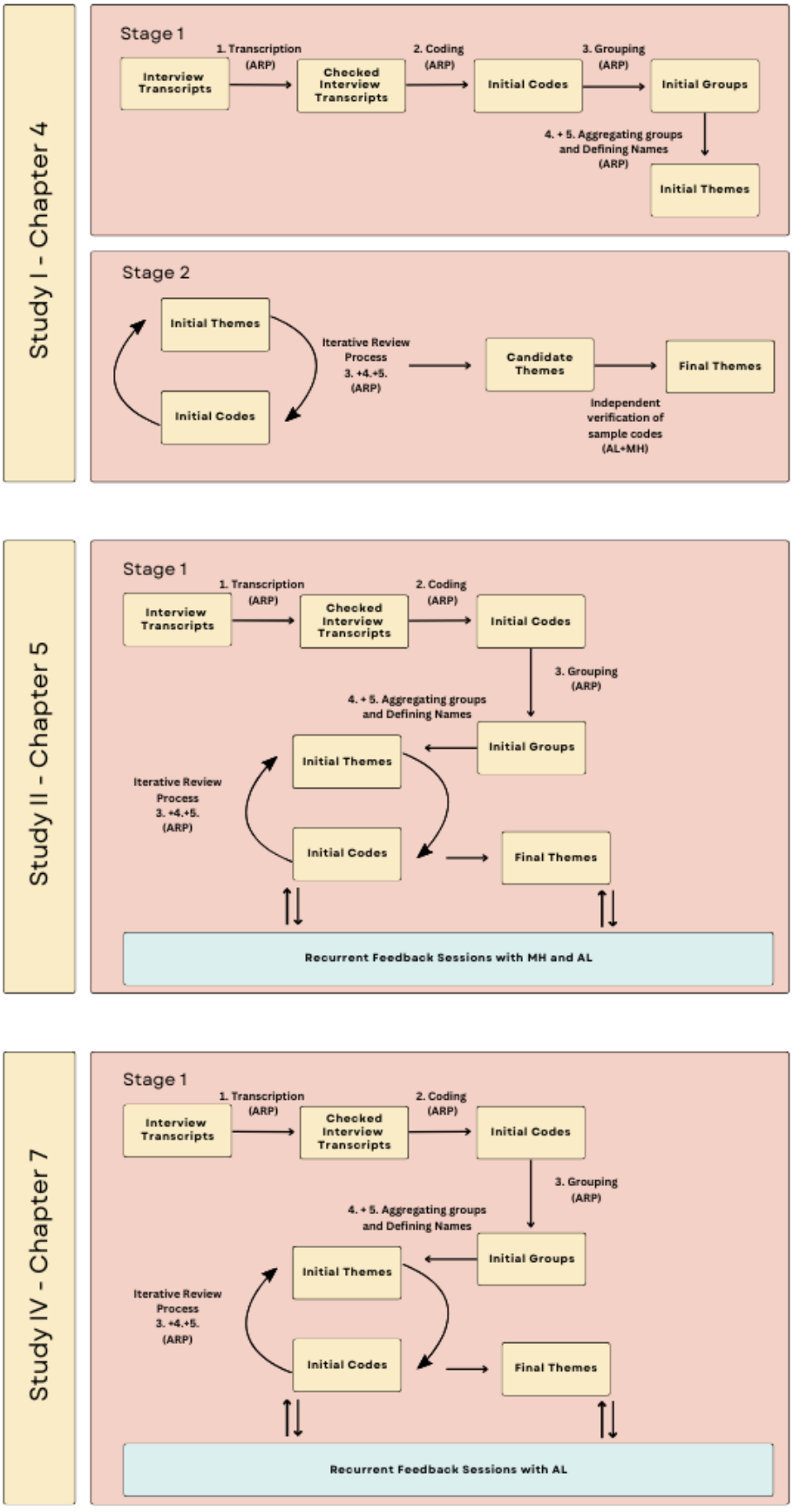


Figure 3.2: Evolution of TA process

3.3 Study Methods

This thesis reports four studies, each with its own methodological configuration based on the stakeholders to be engaged and study aim. This section will describe the methodological choices of each study, to further supplement Chapter 1.5 description of how the studies fit together to form this thesis.

3.3.1 Study I – Experiences and Attitudes towards Consumer Credit Loan Applications – Chapter 4

Aim: The initial consumer interview study aims to understand consumers' experiences applying for loans in the UK Consumer credit and explore the impact of several aspects associated with the automation of this process on their experiences and attitudes towards the process.

What: Study I takes a mixed-method approach combining semi-structured interviews and post-interview online survey.

Why: Semi-structured interviews were chosen as a methodology as they allowed participants to reflect and share their past experiences with the interviewer. As financial circumstances can be sensitive topics one-on-one interviews were chosen over focus groups. Interviews further allow the researcher to ask follow-up questions to discuss chosen topics in a bigger depth or to clarify any misunderstandings, which would not have been possible in a survey and is useful in an initial exploratory study.

The survey enquired participants' attitudes of a list of data sources to be

used on the loan application process. As the list was somewhat extensive, we chose to administer a post-interview survey instead of asking about all the different data sources within the interview as this would be very repetitive and would increase interview fatigue.

As interviews allow the participants to retell experiences based on their own perceptions, they enable the research to empathise and understand the participant's reality [45, 30]. As this thesis has a focus on consumer experience and impact, it was important to try to understand the consumer perspective of the current application processes as a basis to start exploring what the impact would be of a potential DP implementation.

3.3.2 Study II – Consumer Credit Industry Consultation – Chapter 5

Aim: Study II has distinct aims: understand the industry processes associated with the application process and gather some of industry stakeholders' attitudes towards a possible DP implementation.

What: Semi-structured interviews with a variety of stakeholders in industry or in related positions. The interview is divided in two parts each corresponding to a different aim. The first part is a standard online semi-structured interview discussing participants' background in the industry and its working regarding the loan application process. The second part is focused on the potential implementation of DP. As this is a topic which participants do not need to have any previous knowledge there is an initial presentation (Appendix B.4.) by the interviewer (ARP) explaining the technology, where participants can ask any questions to clarify their understanding of DP.

Why: As gaining access to stakeholders from this industry is challenging due to its secrecy and opacity [157] it was deemed more appropriate to answer both research questions within the same research inquiry as they concerned the same set of stakeholders. The method chosen, semi-structured interviews, allowed to answer both research questions. Choosing focus groups instead would not have allowed to answer the first research question due to the private nature of the topics discussed and choosing documentation analysis would not have allowed to answer the second research question as DP has not been implemented in the credit industry in practice. As such, hypothetical scenarios were created and discussed in the second part of the interviews in order to address the second research question. The use of scenarios in the interview has the benefit of giving a bit of distance to the topic for better disclosure.

If the two different research questions had been answered by two separate research activities, RQ2 could have been answered with the analysis of industry documentation as suggested by Tischer et al. [157]. This method could have provided additional information, it could focus on a specific institution or tried to aim to analyse documentation from a range of credit providers. Gaining access to these documents might prove difficult due to the secrecy and opacity of the industry.

The second aim of gathering industry stakeholders' attitudes towards DP implementation could be answered through a focus group with participants from different stakeholders. It could generate discussions based on different stakeholders' aims and needs regarding DP, however participants might also feel less inclined to share specific processes of their institution.

3.3.3 Study III – Differentially Private Decision Based Tree Models – Chapter 6

Aim: The aims of the technical study are to better understand the accuracy drop behaviour of different decision based tree models.

What: The third study consists of a more traditional ML Computer Science study. It is an empirical study where a range of different Differentially Private Decision Based Tree models were implemented on three different datasets. After model implementation performance metrics (AUC-ROC and accuracy) and subgroup accuracy are analysed to better understand the behaviour of the technology within the context studied in the thesis. Chapter 5.1 highlights the model and dataset choices.

Why: As described in Chapter 2.2 Studies on the general public attitudes towards DP tend to discuss the technology in a simplistic manner as if all DP implementations behave similarly, however, this is an oversimplification as seen by the different accuracy drop for the same models with different datasets (seen in Chapter 2 Disparate Accuracy Trade-off). The performance analysis of the models implemented is to some degree rooted on a qualitative perspective, this is, while it uses standard ML quantitative metrics it is done so in a contextual manner (hence the focus on different model and dataset combinations) that focuses on the diversity and variety of accuracy behaviours of the models.

The findings of this study are necessary in a sociotechnical and more qualitative approach to DP, precisely as not to over generalise and simplify the behaviour of the technology. The findings of this study are combined via triangulation in Chapter 8 to understand the impact of DP to consumers.

3.3.4 Study IV – Differentially Private Credit Imaginaries – Chapter 7

Aim: The focus group study aims to gather consumers' attitudes towards the implementation of DP within credit loan applications.

What: This study consists of a focus group involving an interactive game board activity, specifically designed for the activity.

Why: As the technology has not been implemented in this context yet, an interactive game board activity was specifically designed to expose participants to a series of different differentially private credit imaginary scenarios. This further excludes observational studies and document analysis.

The game board component was chosen to more easily visualise and explain the loan application process, and the interrelations between different components while maintaining participant engagement when discussing a very abstract process. The interactive game board activity was a culmination of previous studies as it was designed based on findings from Chapter 4, Chapter 5 and the Literature review (Chapter 2), more details on the design of the game in Chapter 7.2.

As opposed to the first study participants are not asked to discuss their personal financial lives, but are instead discussing imaginary scenarios, as such focus groups are a good method to gather attitudes. Focus groups further allow participants with differing perspectives to discuss it amongst themselves therefore generating rich and detailed data. Had this study been done through one-on-one interviews data collection would have taken longer and the game board activity would have not worked.

"Sociotechnical approaches require a detailed understanding of

dynamic organizational processes, and the occurrence of events over time in addition to knowledge about the intention of actors (situated rationality) and the features of technologies." [145]

3.4 Researcher's Reflections

"In this way, probability is always political" [96]

Throughout my PhD my view and stance on technology, specifically machine learning techniques, has changed and evolved and with it so did my research project. This happened because of its taught component and interdisciplinary nature as well as my personal growth.

My previous educational background has always been in the hard sciences, Physics and Mathematics, therefore from traditional post-positivistic disciplines. However, throughout my educational life I have always had an interest in social science disciplines, specifically gender and queer studies and have also taken Philosophy in secondary school and Philosophy and Politics of Physics in my undergraduate degree. These last modules served to start thinking and realising the impact of social structures within science. This was further expanded in the first year of the PhD by having discussions with colleagues from a variety of backgrounds, reading about ethics of Computer Science and science and technologies studies. As a result, my PhD project also shifted perspectives.

Initially, the aim of my PhD was to combine algorithmic fairness methods with differential privacy to mitigate the impact of the disparate accuracy loss from the privacy-enhancing technology. As I explored the, then-new area, of algorithmic fairness the more I saw the shortcomings of a reductive and discrete approach to such complex problem scenarios. Algorithmic

fairness literature is replete with different fairness metrics based on statistical relations of different components of the confusion matrix, i.e., a table with two rows and two columns that reports the number of true positives, false negatives, false positives, and true negatives of an algorithm's outcomes. This literature tries to simplify complex social situations by distilling fairness into one-dimensional metrics, overlooking the societal and historical mechanisms that have caused these disparities/variations.

Under a more constructivist approach (where I sit on the philosophical orientation spectrum, see figure 3.1) it is impossible to use ML in a fully objective way, as nothing is ever fully objective. It was from this critical view of ML and statistics that I decided to shift my work towards understanding the impact of this technology, DP. To do so it was important considering its context of deployment as well as the attitudes of the end consumers, instead of creating technical "solutions" (thereby stop trying to improve the technology of a system without first questioning the system itself and its underlying dynamics). The shift in the research question and change to a more explicit sociotechnical approach better reflected my personal philosophical positioning (see Figure 3.1).

Due to my educational background being mainly in the physical sciences, I learned more about social sciences and qualitative methodologies throughout the PhD, with Chapter 4 the steepest part of the learning curve. My new position as a social scientist, especially in the earlier stages of the PhD when designing and defining the research activities, probably influenced the conservative choice of methodologies (mainly interviews and focus groups). The last research activity which makes use of the interactive game, is the pinnacle of my evolution as a social researcher so far, showing a greater level of comfort and maturity regarding methodologies and complexity of studies.

Chapter 4

Attitudes and Experiences of Loan Applications: a consumer perspective of the UK context

4.1 Introduction

ML is increasingly being used in a broad range of applications in the financial sector, from front-office, such as automation of Customer Support chats, to back-office tasks such as credit scoring (this study's context), fraud detection and algorithmic trading, etc. [165, 32]. As a result of the Covid-19 pandemic, half of the UK banks see ML and Data Science as becoming more important for future operations [23].

Whilst much is known about these technologies, there is little understanding of how they are experienced by consumers, and how they are perceived regarding fairness in a loan application context. Understanding how users make sense of automated loan applications is an essential step in designing

more transparent, inclusive and widely acceptable processes. Therefore, creating meaningful interactions between lenders and consumers. This in turn could lead to increasing trust in lenders and giving more financial agency to the end consumer.

The main findings of this study are: diversity of consumer preferences regarding modes of interaction and the decision-making process, agreement on the ease of use and accessibility of application process and consumers' desire for more detailed information about how their data was used and explanations of decision outcomes.

4.1.1 Research Aims

This study aims to understand participants' sensemaking of their experiences when applying for loans, as well as their attitudes regarding automation, data sharing and fairness of the process. In this inquiry, automation encompasses processes from the statistical and ML methods used for decision making, to data gathering which makes use of different information systems, to the automation of customer service, as well as the application process itself (for example, short online forms).

The investigation focuses specifically on the UK consumer credit industry. This contribution differs from existing literature regarding algorithmic sensemaking as it addresses the issue of lack of agency on the part of the user in the process. It also provides insight into the lack of consumers' perspective on the role of technology in financial services, identified in Carlsson et al. [34].

Contemporary credit product applications involve interaction with digital technology. This study focuses on consumer experience and perception

of automation, the term 'user' was chosen to refer to applicants when discussing the application process in this chapter to be in line with the literature. Within other fields such as Finance and Marketing the term 'consumer' tends to be used instead, as applicants are purchasing credit products. In this thesis, consumer will be the term used for applicants or previous applicants of consumer credit products.

4.2 Materials and Methods

4.2.1 Research Design

The research study aimed to be an initial in-depth exploration of users' experiences and attitudes. Apart from their experiences with loan applications, the interview guide was designed to explore participants' attitudes toward the process, its automation, data and privacy.

Semi-structured interviews were chosen as a method as they allowed participants to reflect on their past experiences and share their attitudes on the process. This was particularly useful as the thesis is primarily focused on consumer impact, and this study takes an exploratory approach to the consumer perspective of the loan application process. Had the research aims been more based on understanding consumers' actions while applying for credit products a choice of method based on observation would have been a better fit.

The survey was administered after the interview and elicited data about participants' understanding of and attitudes towards specific personal data types, which might be used in automated credit decision-making (both currently being used in the UK and other countries, as well as sources dis-

cussed for potential future use). The interview data was combined with a follow up survey to minimise participants' interview fatigue, as the amount and types of questions on data sources would be very long and repetitive if included in the interview. I chose to have a separate survey that participants could answer in their time via Likert scale-type questions, to be able to still collect this detailed information without massively extending the interview. Surveys based on Likert scale questions have been accepted as standard tools to investigate people´s attitudes [143]. Each scale usually with either 5 or 7 levels, in this research study I chose 5 levels for the scale as for attitudes on data sources 7 levels would provide more granularity than necessary and potentially overwhelm participants. However, there is no standard agreement on what type of data Likert answers are and hence which analysis methods are most adequate [143, 87]. I consider Likert data to be of an ordinal nature and hence in this study using it for descriptive statistics, as opposed to inferential statistics. The survey was chosen to be shared with participants after the interview to allow them time to digest the discussion topics covered in the interview. Furthermore, having the interview before the survey allowed for the clarification and familiarisation with the types of data being asked about in the survey, therefore mitigating the limitations of the Likert survey based on misinterpretation and differing mental models of the questions asked [87].

The study was piloted with an expert qualitative colleague to check the flow of the interview guide and to practice interview moderation, after which the order of some questions were altered. The recruitment strategy, discussed in the next subsection, was subject to feedback from an Equality Diversity and Inclusion expert.

The study design received ethical approval by the School of Computer Science Research Ethics Committee of a UK-based Higher Education institute.

4.2.2 Participants

Recruitment primarily happened via the use of the Call for Participants website, a website providing study recruitment services [1]. Study information was also shared with the researcher's academic networks. As participants were financially remunerated with Amazon vouchers they might feel more inclined to share opinions that they believe might be what I, the researcher, want to hear (confirmation bias). Furthermore, people with positive experiences with loan applications might be more likely to want to part-take in the study. Therefore, there is a limit to the generalisability of the findings based on the data collected.

A total of 25 participants were interviewed and took part in a post-interview follow-up online survey. Participants opted into the interview task and were fully informed that discussions would be about personal loan experiences and perceptions of automation in consumer credit processes.

Participants were eligible if they had applied for a loan with a UK institution in the past, were over the age of 18 and were proficient in English.

A quota sample stratified by tax brackets approach was taken, using the UK tax brackets (in which 0% tax is paid on the first £12,500 earned, 20% is paid on income greater than £12,500 up to £50,000, and 40% is paid on income greater than £50,000, at the time of the study's recruitment). Three other demographic fields were collected, gender, ethnicity, and education level. This data was collected in order to be able to evaluate if the interview sample was representative of the UK population. I have chosen not to analyse by specific subgroups based on demographics as the number of participants for each group would have been too small to comfortably analyse and generalise the results.

The demographics were collected in the form of an open question hence each participant was able to self-identify, in order to try and make the research activity more inclusive [65]. These fields were then aggregated for conciseness and are presented in Table 4.1.

Demographic	n	%
Tax band		
0%	5	20
20%	15	60
40%	4	16
NA	1	4
Gender		
Female	12	48
Male	13	52
Ethnicity		
Black*	13	52
Asian*	4	16
PoC	1	4
White*	7	28
Education		
College*	5	20
Undergraduate	14	56
Degree*		
Postgraduate Degree*	6	24

Table 4.1: Participant's demographic data.

Where Black* was aggregated from : Black, Black American and African American. Asian* was aggregated from : Asian Bangladeshi, British Asian, British South Asian, and Pakistani. White* was aggregated from : White, White British and White other. PoC stands for Person of Colour. Where College* was aggregated from : A levels, College and High School. Undergraduate Degree* was aggregated from : Bachelors degree, Degree, Graduate, Graduate Bachelors, Graduate degree, In university, Undergraduate, Undergraduate Degree, University, University Degree and University Undergraduate. Postgraduate Degree* was aggregated from : Masters, PGC, PhD, Post doctoral degree, Postgraduate.

The ability to self-identify translated to some participants stating their race as black american or african american, terms that tend to be used in the USA but not commonly in the UK. This might be a result of contemporary race discourse being based on the American experience or alternatively, they represent US immigrants in the UK.

The tax band distribution of our sample is representative of the UK's income tax data [79]. There is an almost even split between Female and Males, however there are no gender diverse participants. There is a high level of education among the study participants, possibly as a result of using 'Call for Participants' website for recruitment which might be more visible to students. One could expect for a better understanding of the financial system potentially due to the high literacy level when compared with the general population. The majority of the participants are non-White, hence our participant sample is not representative of the population of England and Wales, which is 86 % White, 3.3 Black and 7.5 Asian according to the data from the 2011 Census [80]. As there is a relation between race and weekly income in the UK [70] this might impact users experiences with the financial system and consequently the data collected as part of this study. According to Francis- Devine [70] Pakistani and Bangladeshi households followed by black households are more affected by income inequality (related by unemployment and wages differences) and have the lowest median weekly household income. More specific data on demographics of loan applicants to different credit products was unavailable.

The distribution of the participants' demographics might be shaped by the recruitment website used. Call for participants does not share the summary demographics of study participants registered with them, therefore it is possible to comment on its representation when compared to the general UK population. Furthermore, as the research activity is financially

remunerated with a £15 Amazon voucher when a larger section of UK residents are unemployed or in furlough during the UK's covid-19 lockdown, that section of the population might be more inclined to participate in the study.

4.2.3 Materials and Procedure

Interviews were recorded online (due to COVID restrictions) over the course of 3 months and lasted between 17 and 42 minutes, averaging a duration of 25 minutes. Interview questions were divided into four sections (Interview Guide in Appendix A2). The first set of questions invited participants to discuss their personal experiences applying for loans. The interviewer (ARP) explained some technical terms that would be used in the interview (e.g. automation, machine learning, privacy) and this then led into the second part of the interview enquiring about attitudes towards automation and data sharing.

After the discussion on data sharing, the interviewer summarised diverse types of personal data sources and asked the participants for their opinions on these being used in loan application decisions. These sources were divided into three main groups: financial data, non-financial and mixed data upon review of the credit lending literature especially based on Hurley and Adebayo [91] , a seminal paper in the literature of the use of new sources of data for credit scoring, and Deville [52], who focuses on the UK context. The classification of data sources was used to summarise family of sources. The initial groups were based on the type of information used in the prime era (financial data) [130] and the non-financial data was based on the new sources of data that have been started to be implemented in the digital subprime era [51, 77]. As some of the data sources found in the

literature such as contextual information regarding financial delinquency or behaviour scores, included both elements of financial and non-financial data, a third category was created to accommodate this. These three categories are then extensive of any possible data sources and have clear names for participants that might have a lower level of financial literacy. For a summary of the data sources used (see Appendix A3 for Interview Guide with summary data sources and A4 for online survey with data sources). Finally, the last section of the interview gathered participants' perceptions about fairness in loan applications and invited them to ideate their ideal process.

Participants responded to a post-interview survey via email, accessing it with a unique code to correlate with interview data. The survey asked participants to rank their acceptance level of diverse types of data being used in automated loan application decisions. The options included data sets that are already used in loan application decision-making in the UK, and other countries, or which are not yet used but implementation is currently being discussed/considered. The survey also included a series of open-ended questions to interrogate combinations of data sources participants were comfortable or uncomfortable with being used, due to associations between different sources. Upon completion of the survey, all participants were reimbursed for their time with a £15 Amazon voucher.

4.2.4 Analysis

The qualitative data elicited was analysed following the traditional 6 step Braun and Clarke [39] TA done in Nvivo, as discussed in Chapter 3.2. As previously mentioned, the analysis of this research study was done in two stages, as a result of my learning experience of qualitative methods. This

first analysis lacked some depth and nuance, which was later reviewed and further built on in Stage 2, after becoming more accustomed to qualitative research work as a result of the PhD and involvement in external research projects.

The survey results depict a finer level of granularity to the attitudes regarding the enquiries about data sharing and was analysed for descriptive statistics as described in the research design section.

The descriptive statistics were analysed at the end of Stage 1 to compared with the data theme developed. Furthermore, interview and survey participant level analysis was done in Stage 2 in the iterative review process, see Chapter 3 Figure 3.2, it was this participant level analysis that led to the development of the semantic scales as a way to summarise the data and trends within it (discussed in more detail in the Discussion section of this chapter).

4.3 Results

The interview analysis produced a headline of themes around:

- participants' experiences with loan applications and the systems used for it (Theme 1)
- the context surrounding the loan application process within the user's life (Theme 2)
- consumer's diverse and complex views regarding perceptions of automation, fairness, and personal preferences of the interactions and decision making of a loan application (Theme 3)

The following section describes the data elicited from the participants, the main themes (T) and core subthemes (SBT) that provide new knowledge on consumer's attitudes towards the current state of loan applications.

Table 4.2. summarises the findings of the study.

Table 4.2.:Theme's Table

Theme	Subtheme	Summary of Findings
T1: General experience and perceptions of current loan application processes and systems.	**SBT1a:** The application process is user friendly	The application process is user friendly and simple to carry out independent of mode of interaction
	SBT1b: Appreciation of Different Modes of Interaction	Face to face interactions when applying for loans allow clarification of terms and empathy but also potential judgement. Online methods allow for a fast process from any physical location.
	SBT1c: Tension between amount of data held by lender and acceptance with use of specific data types	Acceptance of the use of specific types of data if perceived as useful/relevant to the loan application process.
		Discomfort with amount of data shared, caused by feelings of intrusion and security concerns.
		Perceived ideal process for some involves being able to contextualize information and give circumstance and for others to be able to provide less information to others.
	SBT1d: Consumers' levels of understanding levels about the loan process at macro and micro scales	General good understanding of the loan application process across sample
		Knowledge of loan applications and the system is gained by experience and additionally from friends, family and social media.
		Users desire more transparency and understanding about the mechanisms of loan applications and how decisions are made.
T2: Loan applications in the context of a person's life	**SBT2a:** Necessity and urgency of access to credit	Loans are mainly used to cover living expenses and for a minority to gain profits and financial benefits
		Lack of access (partial or total) to loans makes consumers refer to family and friends for financial help
	SBT2b: Emotional impact of application process on user	Negative decision outcomes cause feelings of sadness and demoralisation, which were worsened by lack of understanding.
		Feelings of stress are incurred due to the application process
	SBT2c: Application Process Perceived as non-discriminatory	Application process was not perceived or felt to be discriminatory
		Discrimination was described by users in terms of gender and race but also other demographics and financial status.

Table 4.2.:Theme's Table

T3: Challenges and tensions of diverse user preferences	**SBT3a:** Polarized attitudes towards automation	Diverse attitudes regarding the automation of Loan Applications. Some users consider it as more objective, hence avoiding discrimination whilst others perceive a lack of consideration of people's backgrounds and circumstances.
		Divide between requirement for more automation and speed opposing a desire for more human contact in the application process.
	SBT3b: Differing types of fairness : procedural, contextual and outcome based	Varying definitions and understanding of what fairness in loan application is: mainly procedural, contextual, and based on outcomes and conditions

4.3.1 Theme 1: General experience and perceptions of current loan application processes and systems

Participants generally described their experiences with applications as being user friendly and simple, inclusive of a streamlined application process regardless of online application or in-branch face-to-face application (SBT1a).

> *"**P10** (Male, Black, Und. degree,20 %): Yeah, my experience in applying for the loan was like, really straight forward I did it all online like free in my bank. I just applied online and it was it was really simple like it was almost too easy to apply for it and get it approved. Like, yeah it was really simple."*

However, users had mixed views about different modes of interaction and the function and utility offered to individuals, with a balanced split between the differing groups. Some positively highlighted that face-to-face interactions during the loan application process allows for clarification of misunderstandings, financial terms, back and forth dialogue and empathy.

> *"**P2** (Male, Asian, Und. degree,20 %): I went to the bank and spoke to an adviser. She gave me a bit of information, she explained in some detail, so it made it much easier for me (...) because there's so much financial terminology. Not unless you have a financial background or your you have knowledge of working in the financial sector, you wouldn't know (...) the financial terms are around it's uses."*

In contrast others disclosed concerns about the risk of bias and judgement from human interactions in the loan application process (SBT1b). Where these potential negative factors of in person application processes were expressed, interactions via websites and platforms were preferred.

> *"**P24** (Male, Black, Und. degree,20 %): Humans at the end of the day, they tend to be biased. They might give or they deny you a loan, even if you qualify, but for the machines if the record said you paid you paid the loan if there is no system failure. But if there are no system failures adding to machine so will be good."*

The differences in attitudes towards face-to-face interactions showcase that different sets of users have different needs and hence diversity in modes of interactions from the parts of the lenders might be the best approach to take to be able to meet the different requirements.

Attitudes towards Data Sharing and Survey Results

User's acceptance of data sources and types (financial, non-financial and mixed data types) for use in loan applications is conditional on their perceived usefulness to the process (SBT1c). For example, there was a strong agreement across the participants' sample with the use of financial data as this was seen as necessary for the application:

> *"**P22** (Male, Black, Und. degree,20 %): To start with something like payment income. This is something they need to understand to determine your capacity to pay this loan. They need to have some knowledge on your payment history, whether you*

are financially well or, you are struggling a bit so that they can. They can determine whether you deserve this loan or not. I don't have a problem with that because they are in a business and in a business you have to take associated risks and see how to cover this risk. Defaulting is a very serious scenario and it can make a financial institution go bankrupt, so a lender needs to understand your background, your information, your capacity, and your stability in terms of income and maybe employment, whether you are earning enough to get up for that loan."

These findings resonate with the post-interview survey responses for acceptance level of different types of data sources being used in a loan application (Figure 4.1).

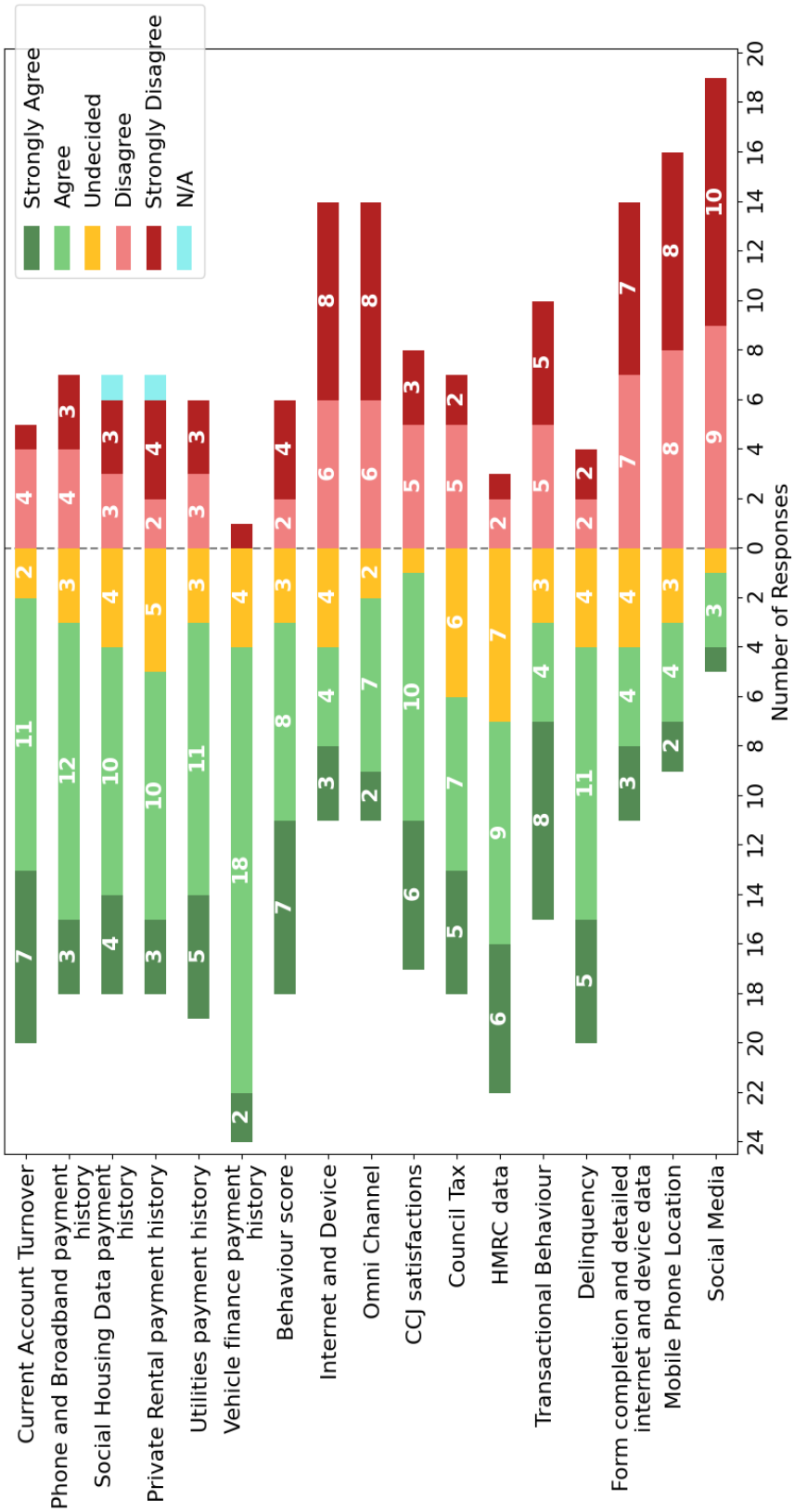


Figure 4.1: Participant's acceptance level with the specified data sources being used for loan application purposes.

Most participants had a positive response (Strongly Agreed and Agreed) with financial data e.g., Vehicle Finance payment history, Current Account Turn Over, being used in this context. Council Tax and transactional behaviour i.e., Behavioural prediction derived from user's transaction history, are the only data sources where there is not a clear majority in terms of agreements or disagreement regarding use in the applications.

For some other types of alternative data sources such as transactional behaviour to provide contextualisation of the financial information (e.g. giving reasons to explain negative events in a person's credit history), there was some agreement, for participants who perceived this information as useful.

The perceived usefulness of this data was subjective and informed by prior personal experience. In reaction to this some users stated a need for increased weighting in the decision making on credit history while others preferred context to contribute more, with these groups not being mutually exclusive (SBT1c). This indicates the variety in user's preferences regarding the application process.

Based on the survey results non-financial data sources, such as social media, mobile phone location and Omni Channels (e.g. phone, email, SMS) among others, are considered as not acceptable to be used in loan applications, by the majority of participants. This is further backed up with interview data :

> *"**P10** (Female, White, Und. degree,20 %): Why do they need like the mobile, my location and stuff like that? That's kind of invasive and doesn't really add much. I would question why that was needed."*

The only non-financial data source that most participants were in favour of

is contextual information on delinquency, as there is then increased visibility and consideration of the individual and their circumstances introduced into the loan application process and decision-making. The contextual information on delinquency (defaulting) would help to understand if it occurred due to external circumstances out of the control of the individual, e.g. loss of a job, or due to"bad" financial decisions.

Participants were also asked about which combinations of data they would or would not feel comfortable with. Participants were mainly opposed to the combination of Omni and Internet information, such as website visit data, due to the potential intrusion to their privacy and the perceived lack of relevance to the loan application process. The survey results support SBT1c in the interview findings.

Interview results disclosed that the aggregate effect of the amount of data held and shared by lenders can cause users discomfort and concerns about data security (SBT1c). Independently of the feelings of discomfort, users were resigned to sharing their personal data, regardless if they trusted the lenders or agreed with the process, due to their lack of agency in (shaping) the process. A minority of participants expressed that they wished they could share a smaller amount of information (SBT1c).

> *"**P12** (Male, PoC, Und. degree,40 %): Okay, I didn't feel comfortable. (...)But again, my hands were tied I really needed the loan, so I really have to just compromise and give that much."*

The general agreement with the use of relevant data sources, associated feelings of intrusion and concern with the amount of data shared and held by lenders leads to conflicting feelings and tensions regarding data sharing.

These feelings of tension showcase the need for a better balance between

gathering, storing and using pertinent data against the risk of intrusion of privacy of the consumer. At this point in time, due to the lack of agency of the users over the process the responsibility of finding this balance inadvertently falls on the industry, and especially the regulator.

Consumers' understanding of the application process

In terms of their understanding of the decision-making process, users have a good understanding at a macro level, i.e. the general criteria used for decision-making and the level of automation in systems used. Most participants were aware of the main factors used in the decision process, such as income and credit history. Participants were also generally aware that current application processes and subsequent decisions tend to be automated (SBT1d).

> *"**P1** (Female, White, Post. degree,0 %): I assume it's like an algorithm where there are certain points attached to each response. Depending on the points value, the loan is awarded or not, that's my assumption."*

Knowledge on the workings of the credit system and loan application processes tended to be distributed via informal networks such as family and friends in the community, as well as online networks making use of social media and websites, or via previous personal experience (SBT1d). Due to the lack of understanding and explanation coming from the industry, there is a missed opportunity for trust building.

> *"**P19** (Male, White, College,20 %): A teacher of mine said the best way to get credit is to get a store card and buy you know one*

or two small things and make sure you pay it off every month and start off with a store card to get your name on the database to get this sort of history for yourself. So eventually when you do need to, when you can apply for a credit card and you do need to apply for a credit card, there will be a history attached to you and it shouldn't be too, too hard."

However, there is evidence that not all consumers are confident in their knowledge of the financial sector and the mechanisms of loan applications, with only half of the participants feeling confident in their current levels of knowledge (SBT1d).

*"**P19** (Male, White, College,20 %): I mean to a certain degree. I'm sure there's loads of other things which I don't. I don't know, loads of things which go right over my head because I'm just not smart enough, you know."*

A subset of participants were not satisfied with the current public information on the process and expressed a desire for more information on the micro level of the process: how data is accessed; how it is used; and the different criteria and weight of each variable (SBT1d).

*"**P7** (Male, White, College, NA): I would like to know the criteria they use (...). Who gives them the information? How they got that check, like when I go to the application, I feed my data which criteria, how, who gives them the authority to check my information you get."*

In addition to requiring more detail on how decisions are made, participants

also expressed a need for explanation for their specific outcomes, especially when they have been rejected (SBT1d).

> *"**P18** (Male, Asian, Post. Degree,20 %): People who don't get it, they should probably get some sort of explanation. I think in terms of, you didn't get it because of people like yourself who have had it had defaulted or people in your salary. People like you who are like you who are working part time in and have been given a loan in the past they have defaulted, or they have had issues, or they have paid it back late. So, in terms of, I mean it would be nice having if you got it as well, but if you got a loan, if I saw that in a piece of paper, I wouldn't read it because it doesn't really matter because I've got it. If I hadn't got the loan and then I want to know why I haven't gotten you and I want to see the reasoning behind it."*

The micro-level detailed information and subsequent explanation of outcomes could provide support to users in making better financial decisions in the future, and help improve consumers' financial knowledge and their confidence in the credit industry.

> *"**P8** (Male, Asian, Post. degree, NA): I mean what is the weight is given to each certain metrics? How much advantages given? So that we (...), as a customer, could really work on those areas to improve it."*

Even when the decision-making is not fully automatised applicants are currently unable to obtain detailed information on the internal decision-making of either a human lender or an automated process due to the lack of

transparency (SBT1d). Participants have an understanding of the factors that are taken into consideration in the process and usually learn it by experience or by shared knowledge in social circles (SBT1d).

4.3.2 Theme 2: Loan application in the context of the consumers' life

The impact of loan products on participants' lives can be deduced by their intended use. Loans are largely obtained to cover living expenses and essential big purchases which include educational costs, home-related costs such as home improvements, replacement of big household items, accommodation costs and more generally to bridge the gap between salary and living expenses (SBT2a).

> *"**P10** (Female, White, Und. Degree, 20 %): I moved into my new house and my bathroom needed basically repairing and I didn't have the money to do it and I needed to get it done like as soon as possible. So, it was for that."*

When participants intended on using the loan to cover living expenses but had their loan application denied or were not given the full amount, they had to resort to borrowing money from friends and family (SBT2a).

> *"**P2** (Male, Asian, Und. Degree, 20 %): So, I found other places, but this wasn't through a loan this was through family and friends who were able to lend it to me."*

A minority of participants described using loans to gain financial benefits, improve their credit score and to invest in the development of their businesses (SBT2a).

> *"**P19** (Male, White, College,20 %): Yeah, I mean I apply for credit cards on quite a regular basis actually. For a few reasons, one because I have really good credit, I get credit cards quite easily, so I normally go through a cash back. Sites like top cash back, they give you good incentives to apply for credit cards. So say for instance you might get in a 50 pound, £400 for applying for a credit card. And then I will use it for the interest free period that it's designed for. I might buy something paid off within it interest free. Then cancel the card and get another card in a years time or a couple of years time and I do this on a regular basis and I believe this is what's helped me keep my credit score very high and very good."*

It is evident from the interviews that participants who described their use of loans for the optimisation of the financial system in their favour appear to have a deeper understanding of finances and credit. This indicates that the current system supports those who are more financially literate possibly as a result of having an informal network that understands the sector (SBT1c), as highlighted by the quote above, and further supports the need for more transparency in the industry.

However, not all users are confident in their financial knowledge. The lack of confidence can impact feelings of agency and control on the part of consumers and might be a factor influencing the user experience and associated feelings of stress when applying for a loan (SBT1b).

> *"**P22** (Male, Black, Und. Degree, 20 %): I perceive it as somehow deceiving. If you don't understand the process and you don't have prior knowledge of what you are doing."*

Independently of the type of loan product or the level of user-friendliness of the application process, the interviews provided evidence that users can experience feelings of stress and worry associated with their interactions of the process, as access to credit had significant impact on participants' lives (SBT2b).

> *"**P3** (Female, Black, Post. Degree, 40 %): At any time you are getting a loan, it is scary, at least for me. I feel very scared. It puts me in a place of anxiety where I feel like, gosh, would I be able to cope, you know, and then you, I'm stuck in this position where I'm thinking about my job security (...) you're thinking about your credit score and how long is going to take your credit score to get back to excellent."*

The lack of understanding of specific detailed information regarding the decision-making especially impacts consumers when denied loans. It causes confusion and users become demoralised by the outcome and their lack of understanding (SBT2b), creating an emotional burden on consumers.

> *"**P6** (Male, White, Post. Degree, 20 %): I came into a situation where I wanted to basically move out, move house and I applied for a loan with them of a higher amount and they actually declined it and I was really confused as to why they had declined it when I had paid back all of the smaller loans on time, so that was really confusing for me."*

> *"**P14** (Female, White, College, 0 %): It really kind of hurt. It's like getting them letters in the post being like you know you've been rejected for this loan, this loan, this loan when I really*

wasn't expecting that, and I couldn't see why. (...), so yeah, definitely quite shocked and a bit hurt."

A small minority of participants felt they had been discriminated against during a previous loan application. Those who had experienced it believed it to be on a basis of income, race and gender which affected the service provided but not access to credit (SBT2c).

*"**P12** (Male, PoC, Und. Degree, 40 %) There was a time I was applying for a loan. (...) I don't know if it's because I was black or something, so I experienced some form of delay or some form of neglect. So, I really had to push hard in the follow up so that I could get the loan."*

This theme showcases the impact that access to credit can have in a consumers lives, it can help deal with unexpected or large expenditures as exemplified in the quote of P10. Some of the findings of this theme also show how much an individual's financial live can cause stress and affect consumers wellbeing.

4.3.3 Theme 3: Challenges and tensions of diverse user preferences

Participants had varied attitudes towards the automation of the loan application (SBT3a). Reflecting on the positive as it made the application process more objective and unbiased due to the lack of emotion and subjective interpretation involved:

*"**P6** (Male, White, Post. Degree ,20 %): See I just think when*

is more computerised there's less of that human element, which is more black and white. It's much clearer."

Others perceived automation in a more negative light as it didn't account for people's circumstances in the application process. The belief being that this could disadvantage users by taking away the human interaction and subsequent human information processing therefore objectifying a person based on a discrete set of variables and lacking empathy.

*"**P3** (Female, Black, Post. Degree, 40 %): It's like computers making assumptions of who I am, it's just like, no you don't know me (...) this is just one facet of my life. But I feel that this facet of my life is being swallowed in by an automated process and then a decision is being made about what my life can look like, what the next step of my life will look like."*

Different views on modes of interaction and levels of automation were also reflected in users' perceptions of the ideal process. Some participants communicated a preference for more automation and speed as these were seen in a positive light - loans could be applied for and potentially be approved quickly and impartially (relating to SBT1b). Others, on the opposing side of the scale, suggested that an ideal process would involve human interaction, and thus allowing for contextualization and empathy (relating to SBT1b).

While the majority of participants sat on each side of the scale regarding automation and face-to-face interactions there were also a minority of participants who advocated for combining the benefits of both approaches. It was suggested this would provide a better overall process, optimising system experience by better meeting the needs of users (SBT3a), specifically

for unusual applications as described in the quote below.

> *"**P18** (Male, Asian, Post. Degree, 20 %): We should perhaps use models (for loan decision-making) but only in certain circumstances. So perhaps we should have humans being part of the process when modelling something which is not usual or different. [...] I think should have more of a human integration . So, for example, if you have an application which is very different circumstances so it's kind of the 5% rare applications. I think that should be looked from a human point of view. But if it's just a common application, which kind of falls into the same kind of the general applications, I think then that can be done by this mathematical modelling, but I do think that it should be a human element involved."*

Overall, there was an acceptance that automation has an important role in loan applications, however, there is also a user requirement for the flexibility provided by a human agent which might add a different dimension to information gathering above and beyond what ML can offer a decision process.

When asked to describe fairness, each participant had a different definition in this context. These definitions appear to relate to their perceived ideal application process. This correlation was identified when cross-referencing codes by participants. These definitions when analysed, where placed in three different groups created as part of the analysis: procedural, contextual and outcome-based fairness (SBT3b). These differing fairness definitions families also reflect different concepts of fairness, mainly the contextual and procedural in the philosophy and legal literature, e.g. the debate between equality of opportunity and equality of outcome, equality and eq-

uity [69].

Procedural fairness is described as the same process being applied to all applicants. It is important to note that the definition of the process itself varies across participants' views, i.e. for some participants the process they referred to was their ideal process, which in itself was very varied and subjective, over defining fairness based on the current loan application process.

> *"**P17** (Female, Black, Post. Grad,20 %): I think fairness is in terms of like the terms of loan application being applied to everyone. And there's not like favouritism and preference for those who access the loans."*

Contextual fairness definitions advocate for adjusting the process according to the applicant's context and needs.

> *"**P14** (Female, White, College, 0 %) I don't think they take a whole story of a person and they just take face value and fairness in this situation would be to look further than face value and understand really what's going on behind a 9-5 job and 9-5 bank account because people have other incomes, people have you know extra jobs or yeah, other people that can help them out for repayments."*

Outcome-based fairness consists of the access to loans being perceived as fair if the applicant is successful :

> *"**P7** (Male, White, College, 0 %): Okay, they are fair [if] they lend you money, that's when they're fair."*

A pattern was identified between participants who define fairness as procedural and those who perceive the current process of loan applications as fair. This reflects the current reality whereby applications and outcomes do involve ML which standardises the process so that it is the same for every individual. In contrast, there is a link between participants who have a contextual view of fairness and those who perceive current loan applications as unfair (SBT3b).

> *"**P3** (Female, Black, Post. Degree, 40 %): I don't think they're fair. I think for you to be successful you have to work for it, you know. And for many years I did not have a credit score and actually had to build a credit score, I had to take out a loan and to get the loan I had higher rates of interest. So, in no way is that fair. I think if it's fair it would take into account different people's situations, and you know I will be able to explain oh I don't have a credit score because I grew up in an economy where there was no credit and I didn't understand it's important to have. You know that that's being fair, that's being just in your process and there's nothing like that. It's the one-way street and you know, choose the highway or get off, you know, so, I'm not sure that it's fair."*

Regarding the impact of automation on the perceived fairness (based on each participant's fairness definition) of the loan application, the interview data was very heterogeneous with no clear subgroups apart from a minority of participants stating that there would be no impact on fairness regardless of if a human or a machine processed the loan application (SBT3b).

4.4 Discussion

This study contributes new knowledge to consumers' views on digitalisation and automation in the financial sector. It explores the perceptions of those who have direct experience with loan applications, with special attention to recruiting participants from varied backgrounds to acknowledge the experiences of underrepresented groups. The study elicits end-consumers' attitudes towards the ideal application process. Consumers are the most affected and often have the least impact on the process design out of all involved stakeholders. Within the algorithmic sensemaking and FACcT literature, to the best of our knowledge, this study is one of the few studies that focuses on a context where consumers have limited agency and interactions with the technology, adding to the field's literature.

The thematic analysis brings to light the heterogeneous views of participants regarding automation and fairness. The diverse standpoints of the participants were varied and rich but with common emergent concepts. A series of semantic scales were constructed to summarise and highlight the individual-based patterns identified but that were not explicitly described in the themes. Semantic scales are usually used as measurement tools for related survey items [72, 50]. In our case these are used as a way to easily visualise related concepts which emerged from the data analysis. The scales feature in Figure 4.2.

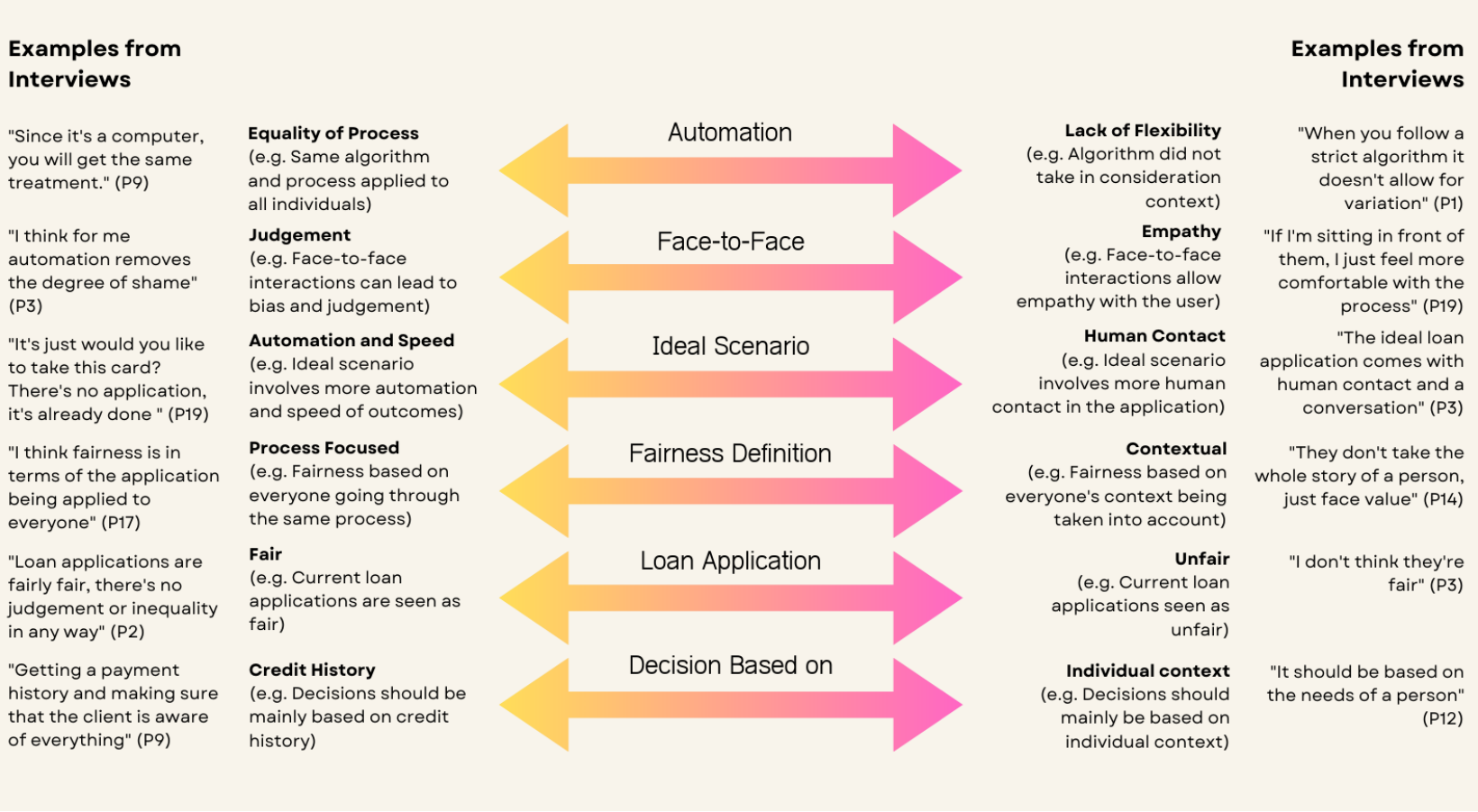


Figure 4.2: Semantic scales on attitudes related to automation and fairness of loan applications. The title on top represents the dimension of the scale and to the left and right we have the extremes of the spectrum and an example quote.

In the semantic scales created, there are two endpoints and a spectrum of opinions in between, as seen in Figure 4.2. The two endpoint scenarios are :

- Ideals of speed and automation- those who see automation as providing objectivity tend to see face-to-face interactions as a source of judgement and are happy with loan applications as their ideals are based on automation and speed. If their fairness definition is based on equality of process they tend to see if as fair due to the uniformity stemming from the automation which is currently mainly based on credit history.
- Ideals of Human Contact and Empathy - those who prefer human contact and face-to-face interactions due its empathy and account of people's context might tend to dislike automation due to its lack of flexibility and see the current loan application process as unfair due to their lack of consideration of people's context in the decision-making.

As semantic scales tend to be used as measurement tools, their design can be based on theoretical frameworks such as the work of O´Quinn [124] or based on initial exploratory research such as in Hallewell et al.´s work [84]. Due to the similarities in approach I decided to follow Hallewell´s procedure for the design of the scales. The scales were developed during the review process of codes in Stage 2 following a similar procedure to Hallewell et al.'s work [84] on user experience design dimensions. Hallewell's process involved two steps:

1. Analyse in detail participants' transcripts
2. Identify dimensions of variance - this stage was iterated based on the research team's feedback

Figure 4.3 shows how the themes based on the diversity of user needs relate to the scales.

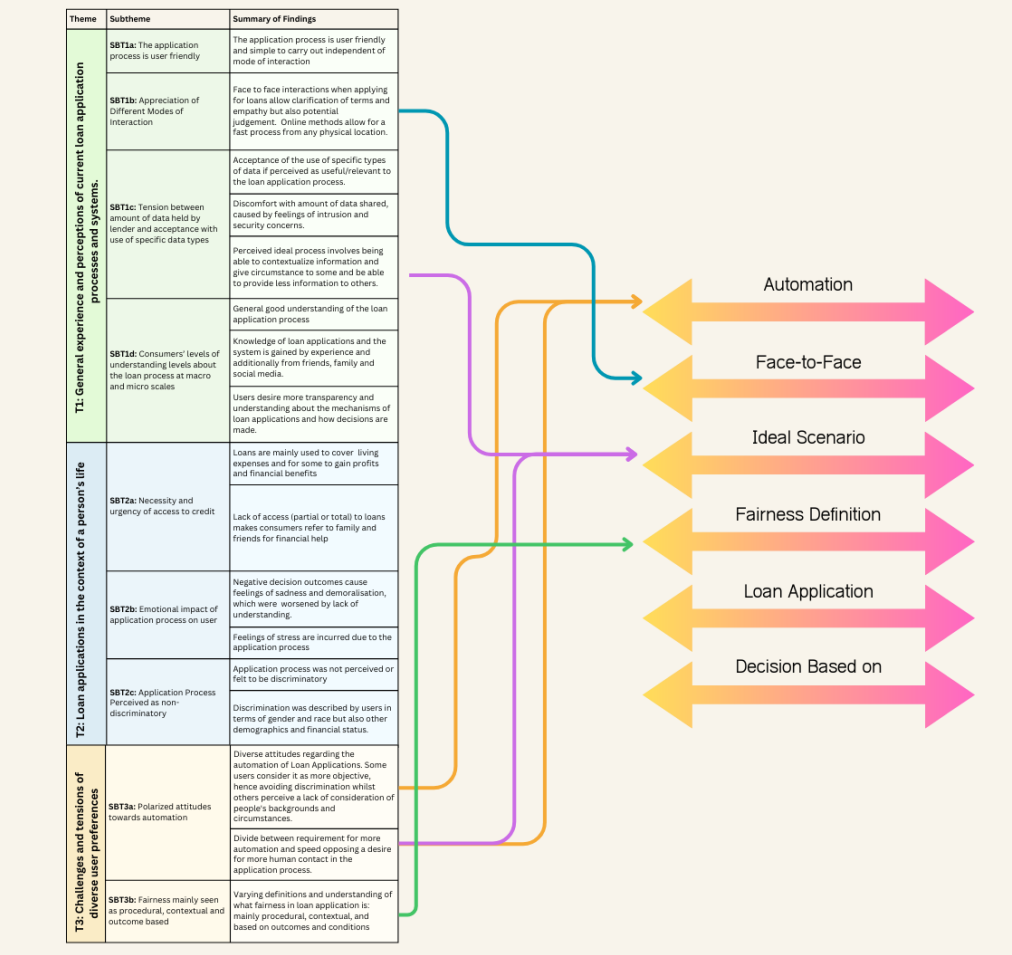


Figure 4.3: Themes' influence on Semantic Scales

While in this study the scales were not tested for reliability or used as measurement tools, in future work they could be the basis of an online survey regarding attitudes towards loan applications in order to understand if there is a statistical significant correlation between the different scales similar to the findings from this interview study.

The findings and the constructed scales fit the subjective and contextual nature of algorithmic fairness perceptions previously published [148, 17, 106]. According to the literature, algorithms tend to be perceived as fair when performing mechanical tasks [17]. The difference between fairness

perceptions in this study might also be explained by participants viewing the loan application process as a mechanical task or a human one.

Bigman and Gray [24] investigated the potential aversion of people to machines making moral decisions (in driving, medicine, military and law fields) and found that people don't want machines making moral decisions as they do not have general intelligence. While decisions in such fields tend to be easily accepted as moral decisions, that might not be the case for decisions in the financial sector. The difference in participant views on fairness definitions and attitudes might relate to their view of it as a moral decision or a purely mechanical one.

The findings further explore topics of the understanding and sensemaking of the application process. According to Xiao et al. [174] knowledge acquired through one's own experiences leaves one better equipped at preventing risky credit behaviour than objective knowledge. Therefore, it is important to increase the transparency of the process, decision-making and design accessible opportunities for people to engage with that information.

Within the algorithmic sensemaking literature, Cotter [42] states users do not know the exact ways in which an algorithm performs tasks but instead have a practical knowledge of how to interact and achieve certain results. This partially resonates with the interview data, whereby users recognised their deficiencies in understanding but wanted more understanding to be able to influence outcomes. Similarly to the process described about loan applications, while not being fully automated user/consumer interactions with the process can seem to be like interactions with complex algorithms. The work of Ironfield-Smith et al. [93] found that despite positive attitudes, consumers required further information about underlying processes, supporting the findings of this study, and identifying specific areas in which

consumers could benefit. Furthermore, the specific lack of information about these decision processes highlights that the efforts to regain consumers' trust by making the process more transparent by part of the industry [4] could be further improved.

Due to the lack of factual user knowledge, i.e. even though participants have a generally good understanding of the process this is not gained through industry explanations and communications. In loan application processes, consumer attitudes regarding automation and fairness are based on their mental models of the process. Meaning that their attitudes regarding the process are not necessarily based on how it happens in practice but instead based on the imagined process (mental model). This mental model is created via their interactions with lenders, their platforms, previous applications as well as informal and formal financial knowledge users might have gained over the course of their lives. The mental models are inclusive of expectations and perceptions of automation and fairness.

This study identifies that the consumer experience of loan applications is multi-faceted, and that improving perceptions of fairness in these processes relies on a delicate balance of designing proportionate and relevant data use in conjunction with both automation and human contact in the decision process. Further work is required to provide insight into how to design such a system, such as co-creation workshops with users as well as technical experts.

Whilst the application process was perceived to be user-friendly, the concept of applying for a loan was still considered stressful due to discomfort with sharing different data types with lenders.

The ease of use of the application process for high-cost loans could be incentivising people to apply unaware of the high fees, similar to the trends

described by Deville et al. [51].

The main intended use of credit products by the participants in this study is for living expenses. This is coherent with the historical trend discussed in Chapter 2.1. of using credit to breach financial gaps, especially during periods of high unemployment such as the Thatcher era [77] and post-2008 global financial crash [74]. The study interviews took place during one of the UK's lockdowns, while a number of people were furloughed or had been made redundant, evidenced by the numbers of vulnerable and low financially resilient consumers [12]. Therefore, these results might vary in differing economic situations and hence further work needs to be done in order to generalise this specific finding.

The emotional burden is further exacerbated by insufficient knowledge of the algorithms that influence how decisions are made, resonating with the FCA's report [12] which found that stress is induced even when the loan application has been successful. To contribute towards alleviating this stress, participants described a perceived ideal process which involves greater transparency of how their information is used, meaning better understanding and visibility of what occurs within the automated decision mechanisms. They also desired explanations regarding why decisions have been made, and having an overall greater understanding of what is happening behind the scenes of their application. These findings further support the call for investment by regulators in the Woolard review [172] towards sustainable and responsible credit.

The setting of this study differs from most other FACcT studies, as the consumer has a lack of agency, i.e., applicants can only choose to apply or not and even this choice is constrained by the need to cover living expenses. Subtheme 1c captures this, whereby participants share discomfort regard-

ing the amount of data shared, yet nevertheless share the information as it is needed to be able to apply for a loan. Furthermore, the desire to understand the outcomes of the applications (SBT1d) is described by some participants as facilitating them to improve their financial behaviour in order to improve their chances in the next application, gaining them more "agency" regarding the application process. This could provide a valuable benefit in terms of user experience if this requirement could be supported by the loan application service.

The findings of this research inquiry have some practical implications for the design of the loan application process and consumer/user-lender interactions. A user-centred design of services based on knowledge of personal MM of these services approach could be taken in the future. Explanations of outcomes and more details on data usage and decision-making could help lenders' efforts to regain consumers' trust due to increased transparency. A better integration of customer service and the application process would help address the main issues experienced by customers (poor customer services, IT system failure and unexpected fees and charges) [12]. Designing services that utilise or combine both modes of interaction, human interaction and ML, would fulfil the need for face-to-face interactions for clarification and iterative communication, with the need for speed of decision-making associated with automation. Integrating impartiality with greater contextualization of an applicant's circumstances would take a bigger change in the industry, however, the study shows a desire for such a process by consumers.

4.4.1 Limitations

As discussed in the research design section of this chapter, there are inherent limitations associated with interviews as a method. It only allows the researcher to gather data from the participant's perspective and due to this research activity being financially remunerated, participants might be inclined to express opinions that they believe might be in line with what the researcher would like to hear, i.e. confirmation bias. In interview studies there is a limited time between the researcher and the participant, when compared with other methods, such as ethnography and consumer pannels with repeated participants, therefore it might be harder for the researcher to create a report with the participants, and not be harder to reach deeper and richer responses from the participants. Potential confirmation bias means that the data collected as part of the interviews might not fully represent the attitudes and perceptions of the consumers.

As this study was my first time leading a qualitative study, it was a constant learning process. Initially, the short interview time combined with my lack of experience might have led to me not have the experience to exhaust a topic. Similarly, with participants that were not very talkative, had I moderated the interviews when I had some more qualitative research experience I might have been able to get participants talking more easily. These issues are common among new interviewers and can be improved with more active listening, better preparation and clearer interview questions, as well as, experience [112]. My interviews skills and how to deal with these challenging aspects improved with each interview. This means that in the first set of interviews certain topics might not have been explored to exhaustion. This however, also occurs with more experienced interviewers, as it is the nature of exploratory research. In the future this could be

mitigated with some more pre-interview testing with participants from the same group that I would be recruiting from.

Despite a recruitment strategy which allowed for time and participant flexibility I was unable to recruit sufficient participants from the higher tax band. Furthermore, within the participants recruited there was a high proportion of non-white participants compared to the general population, which could have some correlations with the tax band distribution, as highlighted previously. There was also a high level of education, which could have stemmed as result of part of the recruitment strategy being based on my personal network as well as using the Call for Participants website, people who are engaged with university research might be aware of. Future studies could recruit a larger and more varied sample to evaluate different proposed models of loan application with varying levels of transparency (etc.). However, it is a difficult topic to recruit due to the sensitivity of the subject, i.e. discussing personal finances, so a more sophisticated recruitment strategy might have to be put in place, such as differing recruitment avenues and methods for different subsections of the population. Future work would consider fully anonymous participation. The sensitivity of the topic might also cause bias in recruitment, as people who have had positive experiences with previous loan applications will be more likely to take part in the study.

While an interpretation is given to the differing opinions regarding automation and fairness of the process, the study does not explore the underlying reasons that might influence participants' attitudes, which could be addressed in further inquiries. Furthermore, the role of agency and perceived agency has been alluded to in the results but requires further examination, including its influence on consumers' attitudes towards loan applications.

Finally, this study focuses on the specific context of the UK and hence the results might not be able to be extended to other socio-cultural settings.

Similarly to the learning process of interview moderation the analysis of this research activity was also my first learning experience of doing thematic analysis. Figure 3.2 from Chapter 3 showcases the evolution of my analysis process. Specifically, in this study the analysis was done in two stages, to guarantee an in-depth and rich analysis. This review generated the themes discussed in this chapter.

4.5 Summary of Findings

To summarise, this research aimed to understand participants' experiences and attitudes with Loan Applications and the role of automation within the process. By conducting interviews and an online survey, we found users' desire for more detailed information about how their data was used and explanations of decision outcomes were described as important for consumers to be able to make better decisions and gain agency and control over their financial lives. The wide variety of consumer preferences regarding modes of interaction and the decision-making process is summarised in the semantic scales designed and would indicate the need for a personalised/bespoke process in most technological contexts. However, within the context of the highly regulated Consumer Credit Industry, the change to a process with varying degrees of contextualisation and automation depending on user and case-by-case application would require a significant change to the assumptions and workings of the industry. The findings of this study are important to understand how consumers experience loan applications and to be able to design services which are inclusive of algorithm-derived out-

puts in a user-centred manner. The results of this study contribute to the underrepresented literature on consumers' experiences with the Consumer Credit Industry.

Chapter 5

UK Consumer Credit Industry: Stakeholder Consultation

5.1 Introduction

While the increase in the use of ML and AI in the financial sector is well documented in the literature [51, 13], see Chapter 2.1.4., there is a lack of in-depth reporting on the factors and the decision-making process behind the implementation of specific technology. The research presented in this chapter addresses this gap and further explores the industry's attitudes towards DP in regards to its implementation in the risk assessment model part of a loan application process.

5.1.1 Research Aims

This interview study with representatives of the UK Consumer Credit Industry developed qualitative evidence on processes directly involved in the loan application of the consumer credit industry. The interview was divided into two parts, each with specific aims and associated research questions:

1. Improve transparency of the Consumer Credit system, including understanding the role and inner workings of the different stakeholders, and interactions between them, in regards to the consumer credit loan application process.

2a. Elicit stakeholders' attitudes towards Differential Privacy

2b. Elicit stakeholders' perspective on potential impacts of its implementation in the risk assessment model in the loan application process.

Aim one is answered in the first part of the interview and aims 2 and 3 on the second.

In the context of the study, the term stakeholder includes lenders (e.g. banks and other financial institutions), credit referencing agencies (CRA), the Financial Conduct Authority (FCA) which is the regulator of the sector, and related third-sector entities.

5.2 Materials and Methods

5.2.1 Research Design

This research study collects data via semi-structured interviews, a type of interview that has a loose structure and a guide but is flexible to the directions participants take, wither direct or not [30], and hence of a qualitative nature. This choice of method allows for extending the survey studies [13] in the literature. By employing interviews instead of surveys we gather data not just on what technology is implemented such as on [13] but data on the decision process behind these choices. This is currently underrepresented in the literature as seen in Chapter 2.1. This method will gain more insight into the industry processes from the perspective of the stakeholders. The interview guide (see Appendix B.3) was iteratively designed (ARP) and independently reviewed (AL and MH).

The guide is divided into two parts: the first part explores the inner workings of the consumer credit industry in regards to consumer credit applications and the second part participant's attitudes towards a Differentially Private Credit Risk Model. For the second part of the interview, participants were also shown a presentation on the workings of DP (see Appendix B.4) to aid with their understanding of the technology. In the final part of the interview, participants were shown three different small hypothetical scenarios as stimulus materials. The scenarios were designed by ARP based on the DP accuracy drop behaviours identified in the literature (see Chapter 2.2.3.) to understand how the accuracy drop behaviour shaped participants' attitudes towards DP.

The study design received ethical approval from the School of Computer Science Research Ethics Committee of the University of Nottingham.

5.2.2 Recruitment

Recruitment followed a snowballing strategy which started with the researcher's professional network (within the credit industry). Snowball sample is a recruitment method in which participants recruited for the study are invited to propose other participants who have experiences relevant to the study [30]. This was chosen as a method due to the difficulty in accessing professionals in the consumer credit industry and the industry's opacity. Recruitment aimed to gather participants from a different range of stakeholders such as lenders, credit reference agencies, regulators, and consumer advocate and policy groups.

Participants were eligible if they worked or had previously worked with or in the UK Consumer Credit Industry, were over the age of 18 and were proficient in English. There was no pre-requisite to be familiar with DP.

5.2.3 Participants

A total of 7 participants were interviewed. Participants opted into the interview task and were fully informed that discussions would be about their personal experiences with the Consumer Credit Industry and attitudes towards Differential Privacy.

Participants had a mix of data science-based roles and business roles with a variety of seniority levels from 6 different institutions. Table 5.1 summarises participant's experience with the industry.

Participant	Job Role Description	Stakeholder
A	Branch manager	Lender
B	Previously Data Scientist and Business Analyst	Lender
C	Chief Data Officer	Lender
D	Previously Non-Executive Director	Regulator
E	Chief Executive	Policy Research Charity
F	General Manager Research and Development	Credit Reference Agency
G	Decision Scientist	Lender

Table 5.1: Participant Summary Table

5.2.4 Materials and Procedure

Interviews were carried out and recorded online over the course of 5 months and lasted between 1 hour and 1 minute and 1 hour and 27 minutes, averaging a duration of 1 hour and 17 minutes. Interview questions were divided into two sections. The first section invited participants to discuss their personal experiences working with(in) the Credit Industry, their day-to-day work, their thoughts on technology use in the industry, etc.

In the second part, the interviewer explained what Differential Privacy was and shared three different technology behaviour scenarios to be discussed, with the aid of a PowerPoint (see Appendix B.4). This part of the interview took between 5 to 10 minutes, depending on the questions asked by the participants. As DP always has a privacy-accuracy trade-off the three scenarios consisted of different accuracy drop behaviours for different training subgroups:

- **Scenario 1:** subgroups with lower accuracy in a non-private model had a bigger decrease compared to those with higher accuracy in a non-private model, hence further increasing accuracy inequality. Behaviour based on Bagdasaryan et al. [15].
- **Scenario 2:** subgroups with high accuracy for the non-private model had a bigger decrease in accuracy compared to groups with low accuracy, hence the private model bridges the accuracy gap between different subgroups. Based on Jaiswal et al.[95].
- **Scenario 3:** there is random accuracy drop behaviour and no clear pattern across groups.

The choice of these three scenarios is based on the DP-SGD literature.

Scenarios 1 and 2 are "extreme" scenarios that could both massively impact different subgroups of consumers, hence good to gather the attitudes of the industry towards them. In most ML real world applications, the technology does not behave in such idealised ways, hence the design of Scenario 3 being the other extreme where there seems to be no rule or trend regarding the accuracy behaviour.

After discussing their views the potential impact of each scenario for consumers, lenders and other institutions, participants are asked to discuss the pros and cons of DP implementation for different stakeholders, and finally to share their general thoughts on the differentially private risk assessment models.

5.2.5 Analysis

The interview data was analysed based on a mix of Thematic analysis based on Braun and Clarke [27] and the Framework Method [73], described in detail in Chapter 3.5.

This research activity was the first time I employed the my personal approach to TA described in Chapter 3, as a result of experience gained in the analysis of Chapter 4 and the associated critical reflection on this process. The analysis of this study was more straightforward.

5.3 Results

Interview analysis produced a series of themes developed around:

- attitudes towards the inner principles of the credit industry and the

usefulness of credit (Theme 1)

- industry balance between lender's drive for profit and regulation focused on consumers (Theme 2)
- the different decisioning components and processes of a loan application (Theme 3)
- the way technology is seen by those in the credit industry and its conservative implementation (Theme 4)
- the varying views regarding DP implementation and the behavioural conditions for this to happen (Theme 5)

The themes and subthemes created are summarised in Table 5.2.

Table 5.2. Themes' Table

Theme	Subtheme
T1: Overall agreement on usefulness of access to credit but diverse views regarding the industry's current workings	**SBT1a:** Agreement on the importance of credit to consumers
	SBT1b: Different agreement levels with inner processes and principles of credit industry
T2: Dynamic balance between lender's drive for profit and regulation focused on consumer impact	**SBT2a:** Credit is a competitive expansive market where profit is a main driver of decisioning
	SBT2b: Data sharing and cooperative approaches not widely used in industry due to sector's competitive nature
	SBT2c: Regulation of the industry on a principle-based approach for consumer protection which puts responsibility on lenders
T3: The loan application process involves different stakeholders and is made up of a credit risk component and a credit policy component	**SBT3a:** Majority or processes in the application based on data provided by CRA, creating a symbiotic relationship with lenders
	SBT3b: Credit Risk Model build involves a back-and-forth conversation with the Model Risk balance to guarantee the right balance of complexity, accuracy among other factors as this is highly impactful for both consumers and lenders
	SBT3c: Credit policy designed based on NPV Modelling which calculates the value of loans to the lenders
	SBT3d: UK the loan application process differs from the rest of the EU due to added affordability checks
T4: Credit Industry generally sees technology as a useful while taking a conservative approach towards its implementation	**SBT4a:** Technology seen as tool that makes processes faster and more accurate however it requires expert personal and good balance between predictive power and complexity
	SBT4b: Lender's technology implementation usually on the conservative side
T5: Implementation of DP unlikely and conditional on the accuracy drop behaviour	**SBT5a:** Measures to ensure Privacy adopted by institutions consisted of restricting access to data and security infrastructure
	SBT5b: While there are pros to DP Implementation, DP unlikely to be implemented
	SBT5c: General agreement that a disparate accuracy drop of different consumer subgroups would not impact them equally but mixed views on preferred scenario
	SBT5d: Varying views and concern regarding the management of the privacy-accuracy trade-off

The following section describes the data elicited from the participants, summarises the main identified themes and the core sub-themes that provide insight into the research enquiry, and new knowledge on the industry's principles and inner processes regarding consumer credit applications as well as attitudes towards DP within the risk assessment model.

T1: Overall agreement on the usefulness of access to credit but diverse views regarding the industry's current workings

Interviews described how credit is seen as useful and important to consumers due to its ability to help spread out costs (SbT1a).

> *"**Pts D:** [Credit] provides the ability of consumers to spread the spending and earnings over time, which they wouldn't be able to otherwise. Essentially what it does is it turns lump sum payments into continual payments over time, and then there's a cost to that which is the interest rates, but yeah, basically allows you to smooth expenditure, which makes sense because usually your income is also smoothed for most people. But expenditure can be lumpy, so what consumer credit basically does is smooth the expenditure so it can be more in line with your income."*

In summary, it allows consumers to afford long-term purchases that they would not be able to do on a short-term basis. As such, the accuracy of decision-making is highly stressed by the majority of participants due to the impact and consequences of an accurate decision. Participant F describes potential consequences of being denied a loan when in need of it to cover living expenses:

> *"**Pts F:** You go to a major lender, and you asked for a loan and you are declined. You still need the money. So, what happen is,*

> *and you can see that in, it's called the debt spiral. You can see people that once they are declined by the major banks they start asking other lenders because they still need the money. But the quality of the lenders start deteriorating very quickly. So, they might go from a major bank to banks which sometimes they have much higher interest rates, sometimes not treating their customers well."*

This is, if a consumer is given a loan they cannot afford this could have negative consequences on their lives. For example, they might default which will be recorded on their financial data and henceforth impact their access to credit products, or alternatively in order to keep up with the repayments of the unaffordable credit, consumers might take out other loans starting to be in a debt spiral. In terms of the impact on lenders, they might lose the money lent out but they might also charge default or late payment fees. If a consumer gets denied a loan they can afford they will lose access to the product and the lenders will lose potential consumers.

The general agreement on the usefulness of access to credit and the extent to which this can impact a person's life discussed here, supports with the findings of Chapter 4 regarding the importance of credit to cover living costs, especially when these are necessary unexpected costs.

The way consumer feedback is taken into consideration within the industry varies a lot from institution to institution, often doing so through the design and compliance departments. The regulator collects consumer feedback through consumer surveys, a consumer panel, and discussions with consumer advocate groups.

Participants had a range of views regarding the industry. The majority

believe it fulfils its purpose and highlighted the potential positive impact of credit. They also highlighted the regulatory role of protecting consumers.

However, not all participants agree with all of the industry's underlying principles and workings. Participants (from the third sector and lenders) were more critical of specific elements of the industry such as:

- types of products
- widespread access to credit
- perceived exploitative use of data for credit risk modelling based on past behaviour to predict future behaviour

In terms of the industry's individualisation of risk and drive for profit (and how these affect the workings of the industry), the main critique of the industry from the data is a disagreement with the balance of individual/-collective responsibility. Evidence from stakeholders provided a view that individual should not bear the price of their credit risk, which leads to risk-based pricing and an increase in inequality. Instead said participants desired a more collective approach to credit risk based on state responsibility, such as increased regulation and investment in the welfare state (SbT1b).

> *"**Pts E:** I think credit can be useful to households. That's fundamentally where I'm from. (...) So, we're all on the same side in terms of advancing credit is generally a good thing to households. Although, my position would be that I'm in favour of dividing up where the market ends and where state responsibility around welfare begins. (...) When we look at actually what's happened around the individualization of risk, then it's*

> *actually led to huge problems for people on low incomes so effectively. When you hook it up to risk-based pricing, you get the poor paying more. And having to bear that all the cost of their own risk, which is not a principal position that I agree with. Although you know if we're in the realm of wanting individuals to be responsible for all their own risk, then that's where we end up. Personally, I think it's a bit better to collectivize some risk at some level. They know to prevent wider societal costs and externalities arising from that."*

Despite the criticism exemplified in the quote above, there was still a desire to change and adapt the industry to a collective risk approach, as credit can have a positive impact on consumers' lives.

T2: Dynamic balance between lender's drive for profit and regulation focused on consumer impact

Most of the criticism described in SbT1b stems from the fact that the credit industry is a competitive market and hence mainly driven by profit (SbT2a).

> *"**Pts E:** This [industry] certainly makes a lot of money. Fundamentally, we're talking about shareholder driven organizations and profit-making enterprises for that purpose. So, it's all about making money for those guys. (...) There's still a prisoner's dilemma for those who want to be the sort of white knight sort of organizations within a competitive marketplace that's driven by shareholder value. They all need the investment. You get the investment by giving better returns to shareholders."*

It is due to profit being the primary motivator for lenders, that the individ-

ualisation of risk and risk-based pricing exists, as it allows lenders to reach a broader population and profit from higher-risk consumers, as described in Chapter 2.1. This type of approach to pricing, however, leads to an increase in inequality as those who can afford the least pay the most, and those that can afford the most pay the least.

Profit considerations can influence all sorts of decisions: from which techniques to employ in the credit risk model, to the design of the credit policy, to passing on operational costs to the consumer, as exemplified in the quote below.

> *"**Pts C:** The primary justification will be a business growth reason. So, either the new technique will unlock more predictive models, which leads to more of those right decisions being made or will lead to kind of more growth of the business, whether that's kind of cost saving or kind of revenue growth, that will be the primary reason. And all within the construct of it being a regulated industry, so you know that's the rules of engagement, we still need to work within that regulation guidelines."*

While consumer credit is a competitive industry it is also a highly regulated one, as described in the quote above by Pts C, which allows the industry to find some type of balance between profit and consumer impact. The regulation of the industry is done through a principle-based approach, which entails not providing specific regulations or guidance but providing a set of principles the industry should abide by (SbT2c). For example, the principle of consumer duty: "A firm must act to deliver good outcomes for retail customers", taken from the FCA's website [10].

> *"**Pts B:** I suppose it would have been related to 2008 like the*

> *global financial crisis and since this overhaul they [regulator] had moved toward a principles-based approach. The idea behind the principles-based approach is to be vague about what the regulator believes is fair or not fair and have the individual companies or industry justify their position."*

The main focus of the principles set out by the regulator is consumer protection, hence for example regulators might not be as interested in which modelling techniques are chosen by lenders but interested in how these affect consumers. What this means is that if DP does not negatively impacts consumers the regulator would not oppose it.

An approach of this kind avoids regulation being exploited through loopholes and puts the responsibility on the industry to justify their choices according to the principles set out in FCA's handbook [10].

The importance of regulation in the industry was exemplified by several participants by the real-case example of Wonga as a cautionary tale of what happens when a company is only focused on profit and ignores regulatory requirements.

> *"**Pts D:** So, for example Wonga, which was a payday loan company was found not to have applied the [creditworthiness] test appropriately to a load of payday loans. (...) If you have given credit inappropriately, and then people can't pay back you shouldn't be able to charge them fees, default fees, but of course people had paid tons of default fees. So, Wonga effectively needed to pay back all these default fees to people that it didn't assess properly. It ended up being too big a sum to pay back, so in fact Wonga went under. A very successful company failed*

because it hadn't done its creditworthiness checks. It made a lot of money up front because it didn't do that. And then it in the end failed because it didn't do what the regulation requires."

Wonga was a subprime lender that had very high-interest rates and became bankrupt as a result of several controversies, due to their unethical behaviours described in the quote above. The Wonga case was one of the drivers to the change in the regulation which put a cap on interest rates [51].

The fact that this industry is of a competitive nature also has implications in terms of the lack of cooperative approaches between firms. While there is a symbiotic relationship between lenders (and other financial institutions) and CRA (credit reference agencies) as will be highlighted in more detail in the next section. Broader data sharing does not happen and is not seen as something beneficial to implement in the future. From the perspective of big banks which have a lot of data, they see data sharing as giving an advantage to competitors without getting anything in return smaller lenders don't have as much data.

Similarly, knowledge exchange of firms individual processes across different institutions in the industry only happens at an informal level, either through personal networks or through participation in industry conferences or surveys and mainly of a technical nature (SbT2b).

*"**Pts G:** In industry conferences we will be fairly open and share the technologies we're using, but then even then like it won't be to the point of sharing what our credit decisioning model looks like, for example. Or here's the exact list of features that are used in the credit decision and moving away from*

> *just the models and thinking more about, (...) you know strategy rules and a lot of other elements that go into making a credit decision, (...) that definitely won't be shared widely outside of the company, so there's almost like different layers of transparency depending on kind of how close you are to the actual decisioning."*

It is therefore the competitive nature of the industry which partly causes its operational opacity. This in turn affects consumer as seen in the findings from Chapter 4.

T3: The loan application process involves different stakeholders and is made up of a credit risk component and a credit policy component

The loan application process is made up of smaller sub processes, with credit risk modelling and credit policy being the main ones. The processes don't just define if an applicant is given a loan but also which credit product, they have access to (different credit limits, interest rates for example).

> *"**Pts G:** If someone applies for a loan, ultimately, that decision of accept or decline is more complicated in terms of what product terms you offer: so that's the interest rates, the loan amount and loan term. But ultimately, that's based on the combination of what the business value is of that particular application combined with the predicted credit risk. But of course they do go hand in hand, but at that point of initial development they're quite separate."*

The risk assessment component of the application is based on the applicant's data. Usually, this data is provided by a Credit Reference Agency

(CRA) and through application forms.

> *"**Pts C:** The models and the credit policies and the decision overlays are all driven by that CRA data so that's kind of the flow of data in the way the UK Industry is constructed."*

Lenders which make use of CRAs then share their consumer's performance back to the bureaus for them to have updated and useful data. Lenders and CRAs have a mutually beneficial relationship (SBT3a).

One of the big steps in the loan application decisioning is the credit risk modelling, which is built to predict the probability of default at a certain point in time (SbT3b) and is used for a range of decisions such as:

> *"**Pts G:** Credit scoring models are the models that are used in decisioning, whether it's for new customer borrowing or for limit credit increases as well as further down the line in calculations like Net Present Value (NPV) calculations, that means what the value is of the borrowing and kind of the program value that lead into wider business decisions."*

The credit risk model-build process tends to be long due to the model governance requirements, which include back-and-forth conversations between the model developers and the Model Risk Office. The continual conversations with the Model Risk Office are part of the lender's internal Governance processes, which across the industry tend to be based on the three lines of defence strategy, defined below:

> *"**Pts G:** [The three lines of defence] it's quite a widely used framework in financial services. The first line is the person itself. So that's someone who's directly involved and responsible*

for building the model and coming up with things like credit decisions and those things. And as part of that, the first lines are also responsible for coming up with the documentation and identifying the risks associated with either the model or the credit decision, or whatever it is that they're working on.

Then the second line is responsible for questioning the first line and validating production, first on a technical basis: that what the first one is doing makes sense. So, if I'm building a new model then second line will check the code that I'm writing actually makes sense. In the case of machine learning model, then second line will ask well, have you selected sensible values for these parameters? So that's on a fairly technical level. The second line also ask questions about if the model makes sense in the changing economic environment. How robust is it? So those are the kind of questions that second line will ask and that's very much independent of first line. Who ultimately owns the risk associated with the model or decision sometimes it's first line, sometimes the second line.

Then the third line is basically an independent audit function, so takes the form of external auditors doing essentially the same job as second line but in an even more independent fashion. So, this will usually take place a lot less regularly (...) That will only happen maybe once a year, once every two years. There will be a third line audit on what, for example, the entire data science team is doing. They're a lot more independent, so it involves a lot more kind of explanation of what you're doing and kind of sharing all documentation you have with them."

This quote highlights the complexity and detail of lender's governance

strategies to comply with the regulator's principle-based approach which puts responsibility and accountability on the lender. The three lines of defense encompasses:

- **First line of defense**: Model developers. Minimize errors, create documentation for models and identify risks.
- **Second line of defense**: Questions first line and validates models. Considers model in wider economic context.
- **Third line of defense**: External and independent auditors, similar tasks to second line.

The implementation of complex models based on machine learning can lead to technical knowledge gaps on the different lines of defense. Participants highlighted the need to upskill the model risk governance for complex models.

A correct loan decision involves providing credit to those who can afford it and denying it to those who cannot, which is highly impactful as discussed in T1, hence the importance of the accuracy of loan decisions and by consequence of credit risk models.

> *"**Pts F:** AI can actually and should be used even more, because at the end of the day, and this is why regulator is very keen, which regulators with new technology tend to be quite conservative. But in this case they were very keen and the simple reason is because AI models can, and they are actually more accurate. If you have a model which is more accurate, it's better for the consumer. But it's also better for the bank because it means I don't loan to people that are not going to be able to repay."*

It is this importance of accuracy that drives new Machine Learning modelling techniques to be employed for the credit risk models.

ML techniques allow for increased predictive power. However, due to their increased flexibility and consequent lack of interpretability they are perceived as an additional risk by some. This is especially true of non-technical stakeholders, hence the need to be more closely monitored.

Also due to the extra flexibility of the modelling techniques, discrimination can seep into the model or due to the nature of the data. Currently, there are different fairness metrics used in different contexts to address this. However, it is still a hard task, especially in the context of credit applications in the UK. The process of credit risk models tends to be built as independently of the credit policy as possible, in order to avoid the propagation of bias between the two sub processes.

The credit policy aspect of a loan application is based on NPV models. These models calculate the values of loans and are used as a base for credit policy definition. NPV stands for Net Present Value and they take the output of the credit risk model and based and assign the value that each risk segment will bring to the company (SbT3b).

Apart from the credit model and the credit policy elements, in the UK, as part of the application process lenders are obliged to do affordability checks which differ from creditworthiness (risk score) (SbT3d).

> *"**Pts D:** Creditworthiness says that you've done a credit check with the credit rating [check] that you use and that the person has come up with a reasonable score. What affordability might require you to do is actually look at the person today and ask them some questions about, you know, have you got the*

income that is gonna cover this? If you're giving someone a mortgage, which is a really big impact on their monthly output, their monthly expenditure. (...) Their past ability to pay back a credit card may not be the best guide, so the credit, I guess the creditworthiness doesn't distinguish by what the actual credit is that they're taking on necessarily. It's a kind of I am creditworthy or I'm not creditworthy as it were, whereas affordability is more targeted, it says is this particular piece of credit affordable. I may have a quite bad credit rating, but nonetheless a particular piece of you knows a relatively sensible arranged overdraft I'm credit worthy for that."

Summarising the quote from Pts D creditworthiness is related to someone's credit score and is based on past financial behaviour. Affordability is not based on credit score and instead looks at the income and expenditure of an individual to see if they can afford to repay the credit according to the terms given by the lender.

While we have gained an understanding of different processes involved in loan application decisions, this process is not transparent to the consumers. This is in part due to several concerns ranging from not wanting the consumer to game the system, to potential negative attitudes of the consumer towards the processes. The small amount of information accessible to consumers about the process is more focused on the technical aspects over credit policy, as this is the intellectual property of the lender. The lack of transparency over the credit policy makes it harder for consumers to adapt their financial lives to have a better chance of getting a loan.

Figure 5.1 summarises the sub processes of the loan application process based on the findings of this theme and T2.

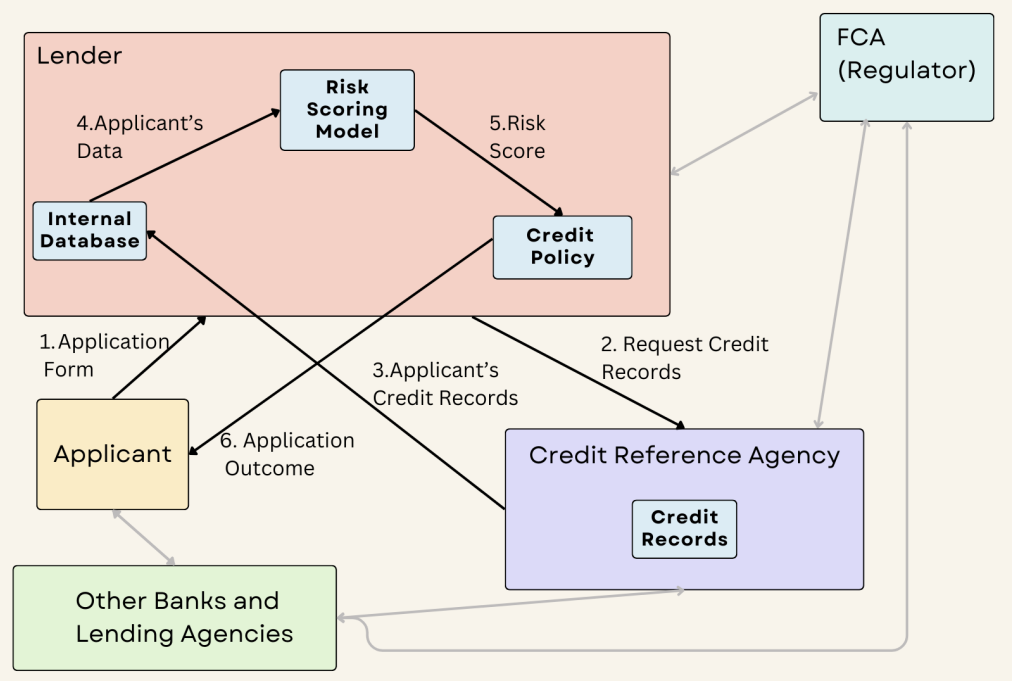


Figure 5.1: Dataflow Chart of the Loan Application Process

Initially an applicant fills a form and submits it to a lender (1). The lender then requests information on the applicant from the CRA (2) who share it with the lender (3). This information is then combined with internal datasets and the form information and is fed into the Risk Scoring Model (4). Out of the model each applicant gets a risk score and this is then combined with the credit policy (5) to see which products and under which conditions the applicant has access to (6). The regulator has some degree of omniscience on all involved institutions (grey unnamed lines) - as they request information and justifications. Furthermore, there is constant information sharing between on financial data lenders and CRA, as well as between lenders and banks and their costumers (grey information lines).

T4: Credit Industry generally sees technology as useful while taking a conservative approach towards its implementation

Technology, and specifically ML, is seen by industry stakeholders as a tool that makes processes faster and more accurate. The improved accuracy is

a result of its increased flexibility and complexity.

> *"**Pts C:** My approach in general is to be very careful where you sit on a complexity versus predictive power spectrum. I think as an organization, we are continually looking for ways to improve the predictiveness of our models. If we get better at that, we make more correct decisions, which leads to better consumer outcomes. And so, we have a responsibility to continually look to improve the predictive power there. However, we have to balance that with the operational risk that might come from an overly complex model. And so, a big focus for us is how do you trade off that complexity against the incremental power that you get from the model?"*

There is a general view that technological implementation needs balance between predictive power and complexity (SbT4a). Models with very high accuracies and high complexity will be hard to monitor and to understand which factors impact its outcome, therefore making them riskier for both lenders and consumers. On the other hand, simple models with low accuracy will make more incorrect decisions therefore potentially causing negative impact to the consumer (potential of having loan products that are not affordable) and to the lenders (loss of profit).

Some participants expressed their belief that the level of complexity of models used in credit risk scoring has reached its peak compared to other more state-of-the-art models, due to the nature of the data used in this context.

> *"**Pts G:** I think in terms of modelling techniques we are getting to a point where we're weighing the trade-off between model per-*

formance and model complexity. It feels like things like Gradient Boosting Machines are probably the best we're going to get for the type of data that we receive in this context so, specifically, think about credit risk scoring context, it seems like there´s not necessarily much sense in progressing beyond to neural networks and introducing all that additional complexity from that because the type of data received is just simple tabular data."

Another participant also raised concerns about the disadvantages of these new technological tools. The main potential negative impact on low-income applicants caused by further segmentation of the market, similarly to the described with the historical introduction of the FICO score in Chapter 2, and the need for trained personnel both for implementation and regulation.

Big lenders are more averse to the use of complex techniques compared to fintechs (usually small companies which leverage technology to provide financial services and products to consumers), but are generally on the more conservative side due to the potential increase in operational risks (SbT4b). Model complexity could also pose a challenge to the regulator as highlighted in the quote below:

*"**Pts C:** Regulators already find it difficult to really understand that models and data that are being used. So, I think that the problem is of a tangible con there from a regulatory perspective."*

Participants views were that compared to other industries which are not as impactful to the lives of consumers and as regulated, the finance industry is not as advanced in ML techniques.

*"**Pts C:** The financial services industry is probably a bit be-*

hind other industries when it comes to these algorithms, if I compare to either retail or more kinds of digital organizations and part of that is the regulatory nature of financial services. I think that there is much more of a human impact for getting decisions wrong compared to potential retail or digital technology organizations."

Within the financial sector, the more complex ML techniques are used in applications different to credit risk assessment, such as for marketing purposes, fraud detection etc. However, the use of ML in these applications in the UK is not as advanced as in other countries such as the US which is seen as a negative point by some as these implementations could improve efficiency.

T5: Implementation of DP unlikely and conditional on the accuracy drop behaviour

The measures currently used in the industry to protect consumer privacy consist of restricting access to certain types of data according to sensitivity and associate level of access, as well as security infrastructure (SbT5a).

*"**Pts B:** The way in which privacy was mainly enforced for individual kinds of applicants or for existing customers is through a system of separation of data. And for the vast majority of people who did have access to the data, they would only ever have access to data which had no personally identifiable pieces of information, so no names, addresses, or anything like that. All of these rows would only be identifiable through a unique number so you join all of the data together just on the application number or just on their customer number. Then in a separate*

> *location, there would be all of the personally identifiable information, and there were more stringent controls around access to that data. It was fairly effective."*

Two out of the seven (A and F) participants had previous knowledge of DP.

After explaining DP, its technicalities and discussing several accuracy drop behaviour scenarios, participants identified both the pros and cons of a possible implementation of DP.

> *"**Pts G:** Of course there are benefits but to achieve these benefits, we're going to make the model predictive performance weaker. So, is that the right thing? (...) But it's more about the level of magnitude I guess. What would you anticipate the drop in performance be and that will be where my concern will be around. Is it a kind of 4th decimal place level of magnitude or is it a significant drop in performance?"*

The positive aspects highlighted by participants include:

- better data protection for consumers
- minimizing reputational risk for lenders
- improved generalizability of credit risk models

The negative aspects discussed include:

- loss of predictive power
- expertise needed (which would involve either hiring personnel or training)
- added complexity to the models

It was hard for participants to quantify the positive aspects of DP, specifically the increase in privacy. While it was possible to quantify the decrease in accuracy which is one of the main negatives as expressed in the quote above by participant G. While DP allows for comparison of the privacy level of different models, the difficulty in translating this to economical terms is a deterrent to its implementation.

When comparing the pros and cons of the technology, the latter are perceived to be more impactful. This could be in part due to the difficulty in quantifying the pros. As such DP is perceived to be unlikely to be implemented (SbT5b). This is in line with the conservative approach to technology implementation already described in T4.

The potential implementation of DP in the Risk Assessment models if there was the drive for such, would depend on the magnitude accuracy drop behaviour and its impact to lender's profit and consumer access to credit.

Industry dynamics are also part of the reason for unlikely deployment, due to the competitive nature of the industry. Individual lenders would not be the first to implement DP as this would mean losing predictive power and hence competitive edge. In turn this means that DP would only be implemented if it was for regulatory reasons.

> *"**Pts B:** Obviously if the FCA start to talk to people about requiring this or we very strongly believe differential privacy is the right way for the industry to be going, because that's the kind of principles-based language that they use, then that conversation would have to happen with the, for instance, the model Risk Office, the independent internal body that matches our risk."*

However, as the decrease in accuracy could negatively affect consumers, and hence go against one of the regulatory priorities, regulators would not likely to support DP. However, this would depend on how DP would impact different consumers.

When discussing the way different accuracy drop behaviours could affect different subgroups of applicants there was a majority agreement that it would not impact people equally (SbT5c).

> *"**Pts C:** If you were to look at a population and focusing on perhaps the population that might be more underserved by credit, and you probably have more, fewer of those people within your data. By adding noise, you kind of wash out some of the accuracy more in that population than others. Equally from the other side, in your more kind of high-end prime population your signal is probably much lower because you have a much lower event rate and again, you kind of run the risk of drowning your signal in very low event rate populations as well. (...) The ability to assess the impact on an independent sample of different approaches, I think that's the only way you can get a clear assessment of exactly how much degradation at what ends of the population you're seeing."*

In summary the quote above states that for people in either extreme of the credit score scale they would have a bigger decrease in accuracy and hence impact compared to the majority of other applicants. A bigger accuracy drop near the low risk segment extremes would be less impactful when compared to risk segments near the lending cut-off point, who might lose access to credit completely.

When reflecting on the scenarios there was no agreement over what type of behaviour within those vignettes would be the most ideal and or the least harmful.

> *"**Pts F:** There will be mistakes because of the accuracy. Now when those mistakes happen as a consumer, do you want to you want to be told because you're in a specific group, you have more or less mistakes or do you want to be told that is a bit of a random thing. Well, I don't know. You could argue that from the consumer point of view, the fairer is to distribute randomly the mistakes but in the overall population, but I could, I mean, I'm having these debates all the time in. That's why I said I can see both ways, but you could also argue that for some groups the mistakes are really impacting much more than in others."*

> *"**Pts B:** With the random one that shape that you've got drawn there is probably the most concerning one of them all from a lending perspective. (...) the most important thing that a model risk officer or a model owner wanted to see is that nice solid line that you've got there with a very kind of monotone relationship between groups."*

As discussed in Theme 1, the loss of accuracy as a result of DP implemen-

tation has the potential to highly impact consumers by wrongly allocating access to credit. Lenders might react by restricting access to credit more broadly, by being more conservative in their credit policies to minimise risk. However, as expressed by the quote below, individual consumers might not realise these impacts due to the opacity of the credit allocation process.

> *"**Pts B:** An individual consumer wouldn't know at all. So, in this case with differential privacy, there's no individual impact because they don't see, they don't ever get asked to prove their income with a bank statement, for instance they didn't know about this, and they won't care about this. But on aggregate it may change how free access to credit is, but again, no individual is able or capable of measuring those kinds of changes. I think it would just be completely outside of most people's experience to even realize the change has happened."*

In terms of how to manage the privacy-accuracy trade–off there were a variety of views. Some stated that lenders should try to equalize the privacy accuracy trade-off and others were concerned with the power this would give lenders hence preferring regulators to advise. It was noted that implementing DP in a different context might have less of an impact as compared to loan applications and one participant suggested that this choice should be given to the applicants (SbT5d).

5.4 Discussion

The aim of this study was multifold: to understand the Consumer Credit Ecosystem, to gather stakeholders' attitudes towards DP implementation

in the risk assessment model of a loan application, and understand potential impacts to the industry of a variety of accuracy drop behaviours caused by DP. In order to achieve this, 7 participants from a range of institutions involved in the consumer credit industry took part in an interview-based study.

The results from the study align with and add more detailed evidence to the current understanding presented in the literature and regulatory reports [172, 130] . Specifically, the general agreement on the usefulness and importance of access to credit and that technology is seen as a powerful tool but implemented in a conservative way in order to manage risk. However, the study adds a level of granularity that is not found in either the literature or regulatory reports, especially regarding:

- the understanding of the processes involved in loan allocation decisioning
- details on the internal governance of lenders
- understanding of the choice of technology to implement for credit risk assessment

As part of the analysis process I realised that the themes developed refer to different levels of processes with the industry and that some of these are related. The themes range from a macro to micro level view of the industry, this is summarised in Figure 5.2. As the DP theme ranges from the micro to macro level and is not as directly related to the rest of the themes on the working of the industry its position within the picture reflects this.

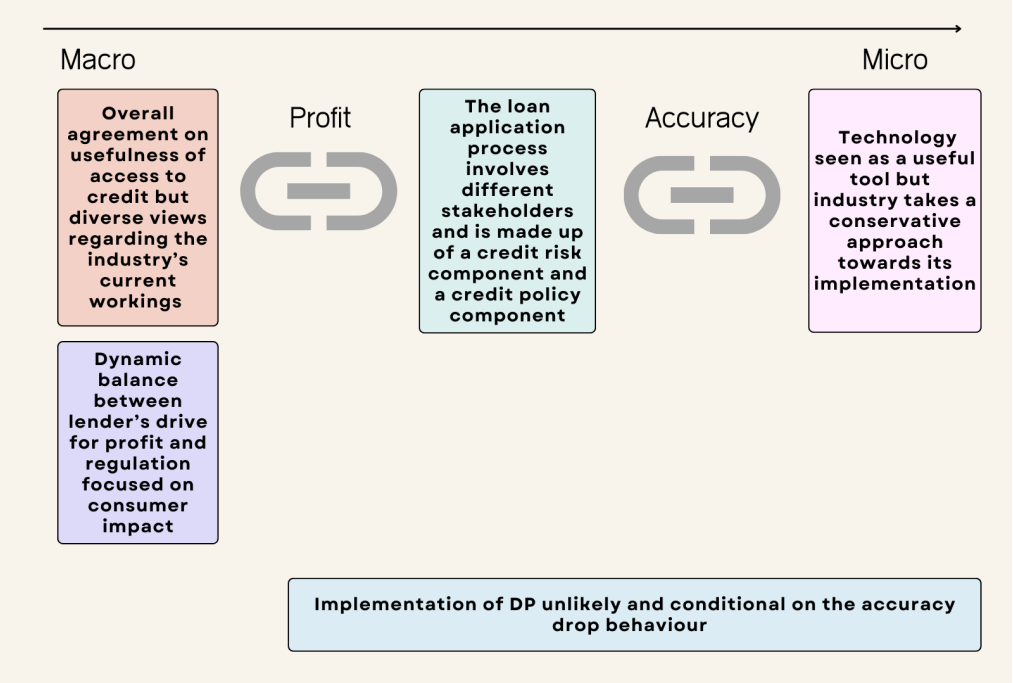


Figure 5.2: Visualisation of Stakeholder Consultation Themes

The concepts of accuracy and profit are related under the current workings of the industry and help relate the different levels of processes. What is meant by this is that the accuracy of the technology models used for credit risk assessment (T4) gets transformed into profit via the credit policy element of the loan application decisioning (T3). At the same time due to it being a competitive industry, profit and by proxy accuracy is high on the values/priorities of most stakeholders (T1). Profit drive is balanced by the regulatory presence and focus on consumer outcome (T2). However, this is not fully accomplished for some subsections of the population due to risk-based pricing, as consumer's socioeconomic background shapes their interactions.

The financial industry is not static as seen by the many changes described in Chapter 2.1., and hence neither is the balance between profit and consumer impact. The Consumer Credit Act was reviewed in 2023 and put into place [159], this balance might suffer some changes in the near future as

institutions adapt to the changes made. Participants had varied levels of agreement about profit being the main driver of the industry and the current balance between profit and consumer duty (T2).

Accuracy and profit are not intrinsically linked (while it will still be important to have good accuracy as one does not want to allocate a loan to applicants who could not afford one) alternatively one could have different credit policies that do not optimise for profit to such a degree, which could mean for example providing credit at a much lower cost to low-income applicants. This subgroup of applicants tend to be riskier and hence tend to pay more for the credit they access. Having a smaller price for credit for this subgroup could translate in fewer defaults and hence better credit history. Furthermore, extending the access to more affordable credit to applicant's that might not be able to afford the current prices of subprime credit.

In terms of the industry views towards DP this is very dependent on the privacy accuracy trade-off (T5). The most important factor is how big the accuracy drop is. If this is small or minimal and hence does not significantly impact consumers it is seen as more likely to be implemented, as compared with a big drop in accuracy which can lead to putting consumers' financial lives in jeopardy. Within the accuracy drop distribution, it is not as impactful if the accuracy drop is bigger on extremes of risk segments as this would not impact loan allocation decisions. However, participants could not agree on whether a random allocation of accuracy drop amongst different subgroups of participants would be better or if it would be preferable to have a larger accuracy drop in subgroups of participants that would not be as negatively impacted by this, e.g. prime customers (low risk costumers which tend to have access to credit with better terms such as lower APRs). These varying views on the least detrimental accuracy drop allo-

cation reflect the subjective nature of fairness and justice allocations, i.e. the common debates of procedural-based justice or impact-based justice in credit loan applications. This debate on procedure vs. impact evaluation is also present in the Discussion Paper and associated industry feedback that the BoE, FCA and PRA have published and hence it is not currently agreed upon or established within the use of AI in the financial sector but might be in the near future [161].

The difficulty in quantifying the impact of better privacy protections versus the drop in accuracy also makes its implementation harder to justify. Another factor influencing the implementation of DP in credit risk assessment models is related to industry dynamics. As there is an associated accuracy drop no lender will want to be the first and/or only to implement DP as it can be seen as a loss of competitiveness. However, if all lenders across the industry implemented DP and had a small accuracy drop there would no longer exist a loss in competitiveness, a first mover problem. Hence for lenders to implement DP it would have to be mandated by the regulator. However, this is unlikely due to the regulator's principle-based approach and due to their role in consumer protection, once again depending on the level of accuracy drop and its impact on the consumers.

Overall DP only seems likely to be implemented if either the general public changes their concept and expectations of privacy to mean more than just data security, which will be explored from the consumer's perspective in Chapter 7 or if there are any data leakage/scandals in the financial sector that might have been prevented using DP, which might make the regulator and/or lenders act.

5.4.1 Limitations

In order to generalise this study and its findings regarding the workings of the industry and their perceptions of DP, it would be beneficial to extend the study to more participants in future work. The sample of participants gathered is varied but not large enough to be able to be representative of the industry. This was caused by the difficulty in recruitment as the Financial Industry is very opaque and difficult to access. Combined with time frame constraints for the study, there was some difficulty to create some rapport with potential participants. There is some sampling bias as the initial recruitment is based on ARP's network. Within the context of the study, the employment of recruitment agencies might not be the best approach due to the opacity of the industry. Upon reflection, a good strategy to improve recruitment would be to attend industry conferences and events in order to bigger industry network for initial recruitment.

Furthermore, had this study been divided into two separate research inquiries, each for each research question, it might have been possible to use different methods. Regarding RQ2 instead of interviewing industry stakeholders one could analyse internal and public documents of the sector as suggested by Tischer et al [157]. This would provide more detail on the processes compared to the data collected. However, it would not be possible to include knowledge on the impact of the social/informal networks in the industry. A mixed method study combining interviews with document analysis could address this point. Gaining access to the internal documents would probably prove quite complicated, as these might contain intellectual property of the firms and hence hard to be granted access to. Regarding RQ3, instead of one-to-one interviews using focus groups could result in interesting discussions and data between different stakeholders on the im-

pact of different models of DP, however participants might not discuss in as much detail aspects related to the working of their institution, in which case an anonymous Delphi method might be suitable if fitting with the time frame.

5.5 Summary of Findings

This study is the first not just to enquire into participants' attitudes towards DP based on different accuracy drops, but also to do so in a specific industry and account for its context. Regarding the current workings of the credit industry, this study shed light on the dynamic balance between profit and expansion and access to credit and consumer protection. Furthermore, it showcased the role that technology currently plays in shaping this industry, by creating new possibilities. These possibilities are then explored and designed e.g. risk scoring and the use of ML, without falling into technological determinism, as the study also showed how the values of the industry shape credit policy which in conjunction with technology shape the industry itself.

Regarding the implementation of DP there is consensus that a disparate accuracy drop of different consumer subgroups would not impact them equally. However, different participants had different views over preferred behaviours. There was also agreement that subgroups in the extremes of risk score would be less impacted overall. The implementation of DP would be dependent on the amount of accuracy loss and regulatory encouragement, therefore currently appears to be unlikely.

Chapter 6

Exploring the effect of DP on Decision Tree based Models applied to Credit Risk Models in Consumer Credit

6.1 Introduction

The Differentially Private Decision Tree based Model study is of an exploratory nature and consists of the implementation of different DP models on three credit-related open-source datasets to compare each algorithm's effect on privacy-accuracy trade-off.

As described in Chapter 2 studies with the general public on their attitudes towards DP tend to discuss the technology in a simplistic manner as if all DP implementations behave similarly. However, this is an oversimplification, as seen by the different accuracy drop for the same models with

different datasets (seen in Chapter 2 Disparate Accuracy Trade-off). This study aims to understand how differrentially private decision based tree models (often used in the credit industry as highlighted in Chapter 2 and Chapter 5) behave in terms of accuracy privacy-trade off as preliminary knowledge. This is vitally important when understanding its impact on consumers.

The findings of this study are necessary in a sociotechnical and qualitative approach to understanding DP, precisely to not over generalise and simplify the behaviour of the technology. The findings of this study are combined via triangulation in Chapter 8, with the findings of the remaining research studies, especially Chapter 5 and 7. Combining the views of industry stakeholders regarding DP implementation with the findings of this chapter, allows one to start to hypothesize and understand the impact of DP to consumers within the application context studied in this thesis.

6.1.1 Recap of Differential Privacy

DP provides a mathematical guarantee of privacy regardless of the potential attacker's computational power and auxiliary data. It guarantees that given a study or query its results will not change considerably if any individual takes part or not. It allows gathering of general information about the population without compromising individual privacy. Differential Privacy can be achieved by the addition of small quantities of noise in a variety of different ways, but in the algorithms implemented in this Study all perturbation happens during the learning process, see Appendix C.

However, DP comes with a privacy-accuracy trade-off, as the addition of

noise reduces the accuracy. This trade-off can be chosen and managed by the privacy parameter (equivalently privacy budget), ε. The amount of noise added to a query depends on the privacy parameter, ε, which can be chosen by us and on the sensitivity of the query, $\Delta\mathcal{M}$. The sensitivity of a query, $\Delta\mathcal{M}$, tells us what is the maximum change possible to the result of the query over all possible neighbouring datasets, i.e. datasets that only differ in one point. Some algorithms use modifications of the sensitivity in order to decrease this value and consequently decrease the amount of noise added and improve the privacy-accuracy trade-off.

6.1.2 Research Aims

The aim of this study is:

- to understand the privacy-accuracy trade-off of differentially private decision-based tree models, for a range of privacy budgets.
- to inquiry the existence of disparate accuracy drops for specific subgroups within the datasets.

It is important to inquiry into the different subgroup accuracy drop as in the applied context in which this thesis is set this can lead to certain subgroups of consumers being differently impacted.

6.2 Experimental Methodology

6.2.1 Datasets

Three different datasets were used across all algorithms tested: a simple Synthetic Dataset generated internally, the Adult dataset and the HELOC dataset. The literature on the privacy-accuracy drop of the DP-SGD (Differentially Private Stochastic Gradient Descent) has shown that there are different results when implementing the same algorithm with datasets with different characteristics [15, 95]. As a result, implementing a variety of datasets with different characteristics allows one to gather results to form a better understanding of the possible scenarios and behaviour of this technology. In the interactive game board study in Chapter 7 we will gather consumer attitudes towards a variety of DP accuracy behaviours which will then be combined with the findings of this study in Chapter 8.

The datasets implemented were chosen to try to optimise the diversity of datasets and computing time necessary to generate the results, as each different model studied was implemented in all three datasets.

Datasets were initially searched in both the UCI repository and in Kaggle and then chosen based on: fit to the problem case, covariate types (had to have both categorical and numerical variables), size of dataset, year of publication and occurrence in literature.

Table 6.1 summarises the main characteristics of the three datasets chosen.

For each dataset chosen, a series of exploratory plots on the correlation between different covariates were created. For covariate correlation, the Cramer's V Correlation Metric was chosen due to the mix of categorical

Dataset	**Synthetic**	**Adult**	**Heloc**
N. of Covariates	2	14	23
Type of Covariates	Categorical, Real	Categorical, Integer	Categorical, Real, Integer, Percentages
Target Variable	default:0,1	income: >50K$,<=50k	Risk Performance: Bad, Good
Number of Records	1000	48842	10460
Year	2022	1996	2016

Table 6.1: Datasets' characteristics

and continuous covariates [131]. Cramer's V can be a biased estimator that can overestimate the strength of the correlation between variables [22]. In this case, as the metric is mainly used as part of an initial exploration of the datasets, these possible biases do not affect the results but can minimally affect the interpretation of these.

The synthetic dataset created has a total of 1000 data points: it consists of default, salary and gender. There is an equal gender split and the salary was defined according to gender: male salaries were drawn from a normal distribution with $\mu = 38000, \sigma = 10000$ and female salaries were ascribed to a fixed value of 20000. This was done to simulate a gender discrepancy in a simple manner and observe the effects of DP. Bagdasaryan et al [14] have shown some disparate accuracy loss in the DP-SGD for gender identification for people with darker skinned faces, however, simplified to try to observe the effect of DP. The target outcome, default, was based on a simple threshold rule: if the salary was smaller than 35000£ then the applicant would default (default=1). The target variable has an unbalanced distribution with a ratio of 7/10 defaults. As expected by the design of this dataset, salary has an exact correlation with default and gender has a strong correlation as well.

The Heloc (Home Equity Line of Credit) dataset is composed of real data. It is a line of credit offered by banks as a percentage of home equity. The dataset holds information on consumers who have applied for this line of credit and it is used to predict if it will be repaid within 2 years, this prediction is then used to decide whether to offer the consumer this line of credit.

The dataset is comprised of 23 covariates which are a mix of categorical and continuous variables. The covariates include fields such as *average months in file*, the *amount of credit products open*, how many months since the oldest and most recent credit product was purchased, and how many products defaulted on at several time intervals among other mainly financial variables. In this dataset, trade represents the purchase of a credit product and delinquency default or missing payments. The target variable of the dataset is Risk Performance which has a balanced split between good and bad performance. From the HELOC dataset, the covariate that has the highest correlation with the target variable is *External Risk Estimate* followed by *Net Fraction Revolving Burden*. This dataset was chosen as it is a more recent dataset compared to the Adult dataset and hence the covariates (type and what they are) are more similar to the datasets used currently in the credit industry today (this knowledge is a result of my experience doing an internship with Capital One UK as part of my PhD program). Hence implementing this dataset will provide findings closer to what would happen in practice if DP were to be implemented.

The Adult dataset was initially extracted by Barry Becker from the 1994 US Census database. The prediction task associated with this dataset is to predict if someone's income is bigger than 50k $ a year (binary), which has an unbalanced outcome. This dataset was chosen even though it is based in the US as it is widely used in publications and has a good size however it is

about income prediction instead of loan applications. It has demographic features and a mixture of 14 categorical and numerical variables. The final weight covariate represents a weighted tally of socio-economic groups where similar demographics will have a similar weight value.

Analysing the correlation from the target variable, based on the Cramers' V metrics, *income*, *Relationship*, *Marital status* and *Capital Gain* are the three most important variables to outcome with *Education* (and *Educational Number*) as well as *Occupation* following.

To summarise the datasets used were the Heloc dataset (balanced target variable), and the Adult dataset (which has an unbalanced target variable), as well as a simpler Synthetic dataset to see how the disparate accuracy loss varies with different datasets.

6.2.2 Performance Metrics

Some of the most common classification prediction metrics used in the literature are accuracy and Area Under ROC Curve (AUC) [134]. Accuracy is the rate between correctly classified cases and the total number of cases, it is widely used as it is not complicated to compute and makes comparisons between different models and datasets straightforward, however, when dealing with unbalanced datasets it is not always useful[88].

Definition 2. Accuracy: The accuracy of a model, binary classification model $\hat{Y} \in \{0, 1\}$, which is evaluated for a dataset with N data points is:

$$\frac{\sum_{i=1}^{N} 1 - |Y(x_i, a_i) - \hat{Y}(x_i, a_i)|}{N}$$

The ROC curve (Receiver operating characteristic) plots the True Positive Rate (true positives overall positives) against the False Negative Rate (false

negatives over all negatives). The AUC gives a single numerical value to represent the Receiver Operating Characteristic (ROC) curve, the closer it is to 1 the better the algorithm is performing, if the value is 0.5 then the model has no separation capacity and finally if it is between 0.5 and 0 it means that the model is classifying it incorrectly.

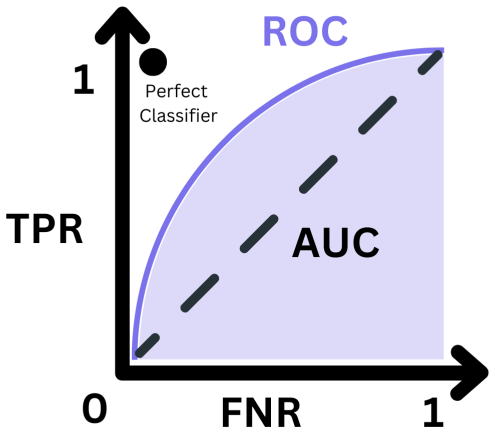


Figure 6.1: Visualization of AUC and ROC

According to Hossin et al. [88] "The AUC was proven theoretically and empirically better than the accuracy metric for evaluating the classifier performance and discriminating an optimal solution during the classification training." There exist other performance metrics such as precision (ratio of correct positives to all positive predictions), recall (ratio of correct positives to all observations in the positive cell), and optimised precision (combines accuracy, sensitivity and specificity), however, these are not as widely used in the literature.

6.2.3 Algorithms and Models

The choice of algorithms to implement and evaluate in this research activity was initially based on selection criteria. Models that were decentralised/federated (i.e., ML technique which trains algorithms via multiple independent devices, each using its own dataset) were excluded as that is not the paradigm of the PhD project. Models which did not have open-access codes were also excluded (due to time restrictions of the study). From that point, models were chosen based on a diversity of methods due to the explorative nature of this study.

Based on the results of [68] the algorithm defined by Fletcher and Islam 2017 Smooth Random Forest SRF, which consists of a Random Forest with Smooth Sensitivity[67] was chosen for the study due to its high utility (i.e. performance) and open-source code.

In order to include more diversity of methods and more recently developed differentially private models, Differentially Privacy Gradient Boosting Decision TreeDPGBDT was selected. DPGBDT is a differentially private GBM which was developed and tested in [107] and is based on the LightGBM open-source library. GBMs are a type of algorithm based on DT which use a different strategy to aggregate/build trees - boosting. These will be discussed in more detail in the Gradient Boosting Decision Tree section below.

A differentially private logistic regression LR algorithm from IBM's Differential Privacy Library was also implemented by me to compare with the DT-based model. In the following subsections, I will briefly discuss each of the algorithms chosen, for more technical details for each of them, please refer to Appendix C.

Furthermore, a completely non-private GBM was implemented, from the commonly used library LightGBM, as a representative of models used in the industry. This model was implemented to be able to have a more "realistic"/"control" comparison.

Smooth Random Forest

The first algorithm implemented makes use of ε- differential privacy and smooth sensitivity[67]. As it is a Random Forest, the process of tree building does not need to query the data, only querying the leaf data for the majority class label making use of the Exponential Mechanism.

The authors were aiming to reduce the amount of noise added by implementing smooth sensitivity over the stricter sensitivity definition.

Each tree in the forest is built without needing to query the data, and the building is stopped when the termination criteria is met, in this case, it is the maximum depth, which is automatically calculated based on theoretical findings by Fan et al. [63].

Each tree is trained on a disjoint dataset, which then uses the full privacy budget allocated to the forest (due to DP parallel composition). This means that the amount of noise added per tree is reduced, when compared to training the forest in non-disjoint datasets.

After empirical tests the authors found the optimal number of trees being between 30 and 100 trees which changes to 100 to 300 trees with bigger privacy budget.

Empirically the Smooth Random Forest Algorithm outperformed a series of other models [94, 71] across a series of different datasets.

Gradient Boosting Decision Tree

There have been several different differentially private implementations of Gradient Boosting Decision Trees [107, 109, 173, 180], however, [107] was chosen over the other models as it performs better.

The GBDT algorithm achieves these results by implementing a novel boosting framework (Ensemble of Ensembles, seen in Figure C.1) and by introducing Gradient-based Data Filtering and Geometric Leaf Clipping to obtain closer bounds on the sensitivity of queries.

Non-private GBDT (same as GBM) train in a gradual, additive and sequential manner so that each new tree improves on the errors of the previous by minimising a loss function with a regularizer term, using gradient descent. Non-private GBDT has a splitting function based on the gradients of the loss function, the Gain of Split. If the current nodes of a single tree have achieved the tree's maximum depth or if the split gain is smaller than zero, tree building is stopped and it becomes a leaf node.

In individual tree construction in the differentially private GBDT implemented, the initial step is Gradient-based Data Filtering which consists of filtering the dataset to be used in training the specific tree. The filtering happens through a simple threshold value of the initial gradient value for each data point, where the threshold is the maximum possible norm gradient in the initialisation.

By performing Gradient-based Data Filtering and Geometric Leaf Clipping the sensitivity of several queries has been reduced, which in turn reduces the amount of noise added when building a single tree.

The Ensemble of Ensemble boosting framework is designed to both make

use of the parallel and sequential composition properties of DP for budget allocation while still maintaining the effectiveness of boosting. Within each Ensemble, trees are trained on disjoint datasets in parallel, and the ensembles are then trained sequentially.

6.2.4 Experiments

All models (excluding LR) were set to have 100 trees total and were trained over 100 different iterations (to average results) for 20 different privacy budgets which were spaced logarithmically between 0 and 10.

The choice of 100 trees is based on the optimal numbers of trees for the Smooth Random Forest algorithm which is between 30 and 100 trees for smaller privacy budgets and 100 to 300 trees with bigger privacy budgets [120]. As we are testing a range of privacy budgets 100 trees were chosen. The number was then maintained for the rest of the algorithms for consistency. The number of iterations, 100, was chosen intuitively to optimise (i.e. minimise) confidence interval with computing time. Finally, all models also had 100 different iterations of 100 trees for a Non-private limit (by setting the privacy budget to 1,000,000,000, as this number is large enough to represent infinity in a quantifiable computable manner). For the LR model, we trained 1000 iterations.

The privacy parameter were logarithmically spaced due to the behaviour of the privacy-accuracy trade off curves (see Figure 1.4)- where the biggest change in values is for smaller values which tend to follow a logarithmic like trend.

In practical applications of DP there seems to be no clear consensus of how to choose privacy budget, ε, with different companies differing in budgets

by orders of magnitude [56], with some companies having a budget of 0.1 and others of 4. Within the academic community Ganev et al. [75] test budgets from 0.01 to 100 but most other research is within the ranges from 3 to 7 [15, 177] when ε is disclosed. In this study we drew 20 different privacy budgets logarithmically spaced ranging from 0.1 to 10. This means that the majority of the budgets are small, and this was chosen because of the theoretical asymptomatic behavior of the privacy-accuracy trade-off graphs, such as the one in Figure 1.4.

For each privacy budget the $\hat{Y}$ was written into a file for all 100 iterations, furthermore, after the model has been through all 20 budgets the Accuracy and AUC metrics for each were written into a file. Outputs were similar for the non-private limit.

Hyperparameters were not optimised for any of the different models and were kept the same across different methods as much as possible.

LightGBM, a commonly used library of GBM models was also implemented in order to compare with a fully non-private algorithm, as even in the Non-private limits of the Private models the structure of the algorithms themselves are still optimised for DP. The implementation of the LightGBM allows one to have a more realistic and representative view of Industry practices when implementing new models.

All the code was run in a University server over the course of 5 months, after an initial year of testing all models and other alternatives on a smaller scale.

6.3 Results

6.3.1 Overall Performance across Privacy Budgets

In the context of this thesis' work, i.e. applied context focused on the credit industry, it is important to compare models based on performance as this is the basis for the choice of implementation (see Chapter 5.4.), the higher the accuracy of a model the better the chances of it being implemented, as it can lead to better and more specific loan allocation, leading to increase profitability of the lender. Better accuracy of models also positively impacts consumers as they will be allocated credit products which they have the capacity to repay.

Figure 6.2 shows the Accuracy and ROC of different model implementations with the different datasets respectively for all the privacy budgets (x-axis), as well as their non-private limits (point dash lines). The black line corresponds to the accuracy of the LightGBM implementation.

The ideal performance plot has two elements: a very steep "climb" as close to zero Privacy Budget (x-axis) and a "plateau" close to 1 (or close to the non-private limit). A graph like this means that even at very small privacy budgets, same as high privacy guarantees, the model would maintain a good performance. On the other side of the spectrum, a differentially private model that is not a good performer will have a low overall accuracy and climb which tends to start in the higher Privacy Budgets, hence weak privacy guarantees.

Based on the findings shown in Figure 6.2 we can observe:

- **Different models performance differs dependent on the num-**

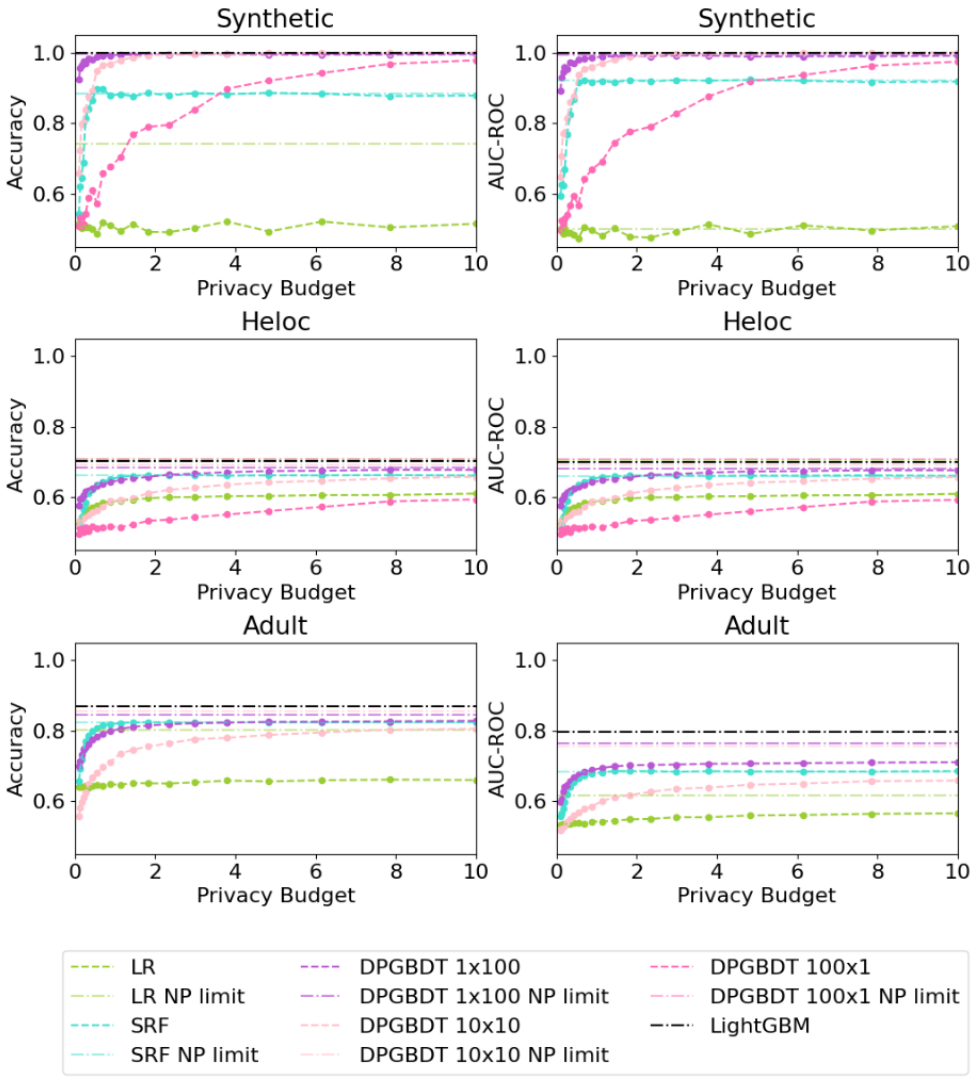


Figure 6.2: Model comparison of accuracy and ROC for each dataset

ber of categorical variables. Comparing across datasets LR model does not perform well in datasets that have a bigger percentage of categorical variables (Synthetic and Adult) when compared to datasets with more numerical variables (Heloc). The opposite appears to be true for DPGBDT 100x1. For the remaining models, there doesn't appear to be an impact on performance.

- **DPGBDT always reaches the non-private limit** (curves for this models in Figure 6.2 reach the straight line). For the Heloc dataset and the 1x100 architecture model even surpasses the non-

private model accuracy and ROC.

- **DPGBDT 1x100 has the best starting performance metric.** (starting point for this model in Figure 6.2 is higher than the rest of the models).
- **SRF and DPGBDT 1x100 both have equally steep climbs.**

What does this mean in practice: In practise DPGBDT 1x100 would be the model chosen by the industry across all models tested as this one is the one that consistently performs well, both in terms of overall accuracy as well as in terms of steep climb. This means that in practise if the model were to be deployed it would be possible to have a small privacy budget (which equates to high privacy guarantees) and still maintain an overall accuracy within the same range as a non-private model (here exemplified by LightGBM).

For each model and dataset combination, I further identified the best accuracy-privacy trade-off point by finding the elbow point, i.e. the point in the climb where the performance metric starts to plateau (using the *kneed* python library) in each of the graphs of figure 6.2.

Figure 6.3 plots the points in the same axis. One of the benefits of the technical language of DP is the possibility of quantifying privacy in terms of model comparison, i.e. one can equate privacy levels of different models given that they have the same privacy budget, ε.

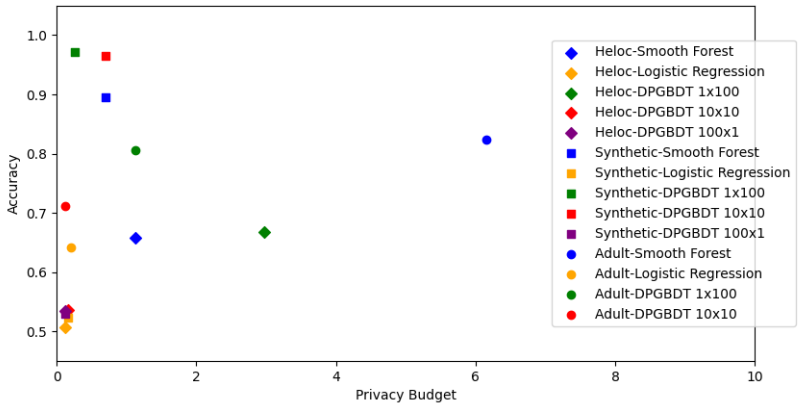


Figure 6.3: Comparison of best privacy-accuracy-privacy trade-off point across models

By focusing on points with the same colour, the Figure confirms that **generally considering all datasets the DPGBDT 1x100 is the best-performing model**, as all points are the closest to the left-hand corner of the graph which is the ideal behaviour, with Smooth Forest as a second runner, even outperforming DPGBDT 1x100 for the Heloc dataset.

6.3.2 Subgroup Accuracy Behaviour

In this section, we analyse the accuracy values for the different subgroups or possible values taken for each covariate. In order to do so accuracy plots have been generated for each covariate for 3 different values of privacy budget, i.e. $\varepsilon = 0.162, 1.44, 3.79$ and the non-private limit. These specific values were chosen as I wanted to include a very small value (0.162), a value around the peak of the privacy accuracy trade-off curve (1.44), and a value close to non-private limit in the privacy-accuracy curve (3.79 as general accuracy similar to accuracy with a budget of 10).

In line with Xu et al. [176], we consider accuracy loss disparate for subgroups if this decrease is of a different order of magnitude of other sub-

groups, i.e. meaning at least a 10x bigger decrease in accuracy. Furthermore, we consider ODAL (Opposite Disparate Accuracy Loss) if a specific value of a covariate has at least a 10x less decrease in accuracy.

Figure 6.4 graphically summaries the covariates for each there is at least one occurrence of DAL or ODAL for all three privacy parameters and for each of the subgroups (same as possible covariate values) of each covariate.

Figure 6.4: Presence of DAL or ODAL for each covariate by dataset and model combination.

What the figure indicates is that the occurrence of both DAL and ODAL is sparse, especially DAL.

The model which has more occurrences is DPGBDT 1x100 with an equal

amount of DAL and ODAL.

There is ODAL for Net Fraction Revolving Burden 15-40 , see Figure 6.5.

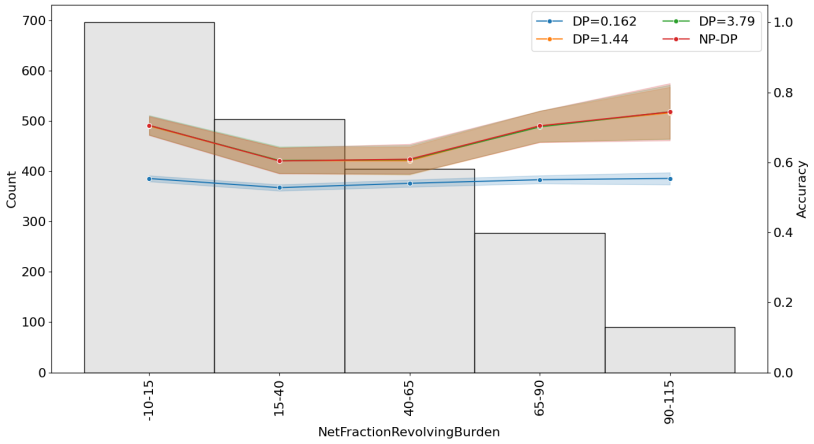


Figure 6.5: DPGBDT 1x100: Net Fraction Revolving Balance Covariate Accuracy

Often both cases of DAL and ODAL, disparity stems from the cases where the privacy parameter is so low that its accuracy is around 0.5 for all subgroups but that for higher values of privacy parameter and even for the non-private limit subgroup performance varies, such as shown above.

This explains why the two best performing models (DPGBDT 1x100 and SRF) tend to have the highest amounts of DAL and ODAL.

When this is not the case the covariate graphs are similar to Figure 6.6

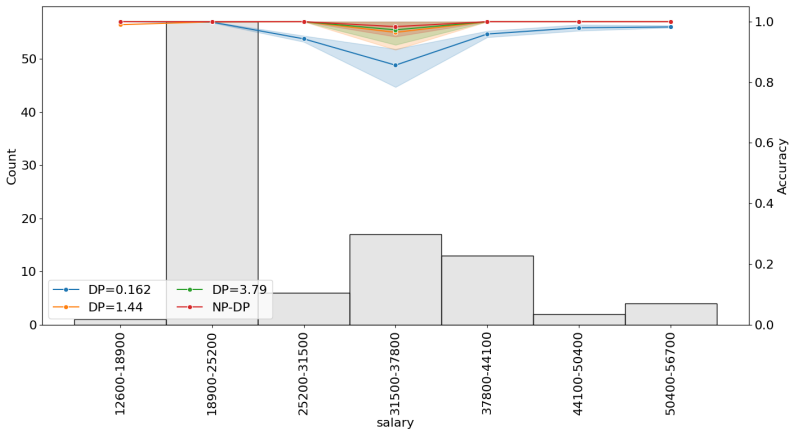


Figure 6.6: DPGBDT 1x100: Salary Covariate Accuracy

Furthermore, the differences in changes in accuracies for the different models are related to the "speed" of climb in Figure 6.2, as even though we are comparing the models on the same privacy budgets, the corresponding overall accuracies vary.

What this means in practice: As a result of the reduced occurrence of both DAL and ODAL the impact to the accuracy of the model for implementing DP would probably not affect any subgroup disparately (in terms of accuracy of model- how that translates to wider impact will be discussed in more depth in both Chapter 7 and 8).

6.4 Discussion

DPGBDT 1x100 stands out as the best-performing model across the datasets, retaining patterns and accuracy behaviour even in very small privacy budgets $\varepsilon = 0.162$.

The SRF model performs equally well for privacy budgets between $\varepsilon \in [1.44, 3.79]$ as shown by the accuracy and the non-private limit being nearly

identical, however in the lowest privacy budget it does not perform as well as DPGBDT 1x100.

There is a broader range in model performance for the Synthetic dataset, followed by the Adult and finally the Heloc dataset. This indicates that models handle numerical variables similarly well, however, the same cannot be said about categorical variables.

Within DP literature the privacy budgets tested range from 3 to 7 [14, 176], with the exception of Ganev et al. [75] which considers ranges from 0.01 to 100.

This work compares from 0.162 to 3.79. This work adds a finer level of examination for the small privacy budgets. This is especially important as we are taking a user-centred approach and as such one would aim to have the bigger amount of privacy possible given performance constraints.

Apart from LR, the models implemented had a similar if not better performance compared to DPSGD on the Adult dataset for $\varepsilon = 3.1$ implemented in Xu et al. [176] and similar subgroup accuracy behaviours.

Findings in Xu et al.'s work state that the DAL from the DPSGD stem from the gradient filtering. As DPGBDT also has two types of filtering one could expect a larger number of DAL, which is not the case. This could be a result of the Ensemble of Ensemble's structure, which differs from the sequential structure of DPSGD.

Overall after analysing all datasets the impact in terms of disparate accuracy drop, DAL, over all datasets is minimal and often models have a subgroup which decreases significantly less than all others, ODAL.

This behaviour can be easily understood from a theoretical perspective by

looking at the two extreme case limits. If we had a perfect non-private model, it would have accuracy and ROC of 1, and consequently, all subgroups would also have perfect performance metrics. On the other hand, if we had infinite/perfect privacy then our accuracy and ROC would be 0.5 (for a balanced dataset), for binary classification tasks (as all predictions would be the same). However, when we compare a non-perfect model with variations in subgroup performance (as most models have) with a low privacy budget model where subgroups' accuracy tends to 0.5 we get DAL, which is what happens in a lot of the disparate accuracy drops identified for the combination of datasets and combinations implemented in this study.

What this indicates to us is that in order to avoid a disparate impact in subgroups, choosing a model that has a steep climb and a good accuracy even for small values of ε, so that we can set our privacy budget lower and maintain the subgroup accuracy patterns while preserving privacy at a higher level and maintain a good performance.

Regarding its impact on the industry, while the results in this study are not based on datasets that are currently used in the UK Consumer Credit History, the Heloc dataset is real-life credit data whose results seem positive in the sense that for a privacy budget above 4 (for the DPGBDT 1x100) the accuracy is only lower to the LightGBM's by 0.02- 0.03. If the DPGBDT 1x100 model has the same behaviour in a real-life dataset then, at least on an aggregate level, its impact on the loan application process could be deemed acceptable by the Industry and some Users (see Chapter Industry Study, Chapter Focus Group). Chapter 8 explores how the findings of this study, combined with the rest of the findings answer the Overall Research Question.

6.4.1 Limitations

Some of the limitations of this work include the lack of theoretical analysis of the algorithms implemented, and the non-optimisation of hyperparameters (which would lead to an increase in overall accuracy) and would happen in a practical implementation of DP, hence the findings can only be extrapolated to the consumer credit industry to a certain level. The same can be said about the datasets, while Heloc as a dataset is fairly similar to datasets used in practice, this is based on USA data, furthermore different lending companies will not all have the same datasets, depending on the CRAs used and their own internal data.

The choices of datasets and parameters in this study were based on the public knowledge regarding the industry and their processes as well as findings from Chapter 5, this is however limited due to the high opacity of this industry. From Chapter 5, we know that DP has been implemented and tested in the industry in different scenarios other than risk assessment models but these findings and information were not made public (even aggregated general findings). If Lenders and other financial private institutions shared their RD more openly without disclosing proprietary IP, researching in and about the industry would be more accessible and would help decrease its opacity to consumers, as well as making the understanding the impact of differing technological implementations such as DP easier to investigate.

Furthermore, at the point of writing this is the first publicly available study enquiring into the potential existence of DAL in differentially private decision tree-based models, and in order to generalise it and further validate these findings these models should be implemented with a wider range of different types of datasets.

6.5 Summary of Findings

The Differentially Private Tree-based Model study is the first study on the comparison of performance for a variety of decision tree-based models. The study is of an exploratory nature and consists of the implementation of different DP models on three credit-related open-source datasets to compare each algorithm's effect on accuracy and subgroup accuracy.

Key Findings to Overall Research Question:

- Over all models and datasets, the occurrence of DAL (Disparate Accuracy Loss) is minimal but ODAL (Oppposite Disparate Accuracy Loss) are more common
- DPGBDT 1x100 is the best-performing model across the ones evaluated when considering all datasets
- DPGBDT 1x100 for privacy budgets $\varepsilon > 4$ has an accuracy comparable to the non-private model implemented (LightGBM)

These findings of this study are necessary as not to over generalise and simplify the behaviour of the technology. The findings of this study are combined via triangulation in Chapter 8, with the findings of the remaining research studies, especially Chapter 5 and 7. Combining the perspective of industry stakeholders with the findings of this chapter allows one to start to hypothesizing and understanding the impact of DP to consumers within the application context studied in this thesis.

Chapter 7

Consumer's Exploration of Differentially Private Sociotechnical Credit Imaginaries

7.1 Introduction

The Differentially Private Consumer Credit Imaginaries Study consists of an interactive focus group activity which aimed to understand how consumers perceive the implementation of Differential Privacy in different scenarios (RQ5). These different scenarios were communicated via a 'game' which was designed specifically for this research enquiry as a multi-purpose tool. The game board serves both as an educational tool, to ensure a consistent baseline understanding of DP by all participants, as well as to prompt focus group discussion about DP scenarios.

This study is the culmination of new knowledge and understanding of the credit industry. The experiences shared by participants from Chapter 4 helped design the personas of the game. The findings on knowledge of the loan application helped establish which type of language to use in the game. Furthermore, this knowledge informed which information could be shared as part of the game and which required more attention and hence should be preemptively in the initial presentation to aid understanding of abstract concepts. The findings regarding the loan application process and its sub processes, from Chapter 5, helped design the structure of the board. The understanding of the relation between the different sub processes helped design the play dynamics. The findings on the preferences of the different DP behaviours explored in Chapter 5 served as a basis to design the DP models of the game board. The knowledge on how DP might affect the consumer credit industry was used by me as the moderator of the activity to further question the participants by playing the role of the industry view.

7.1.1 Research Aims

The study involves in-person game-based interactive focus group to discuss participant's attitudes towards Differential Privacy in Credit. For this study participants did not have any background knowledge on DP.

The study seeks to understand how different accuracy drop behaviours affect participants' attitudes towards DP implementation. Furthermore, the study was designed to expose and gather participants' views on the processes and stakeholders associated with the loan application process, via the game board activity.

7.2 Materials and Methods

7.2.1 Research Design

The methodology chosen for this study consisted of a focus group interview. This was facilitated by the integration of a bespoke 'game', henceforward referred to as DP loan game. It was designed to create an interactive and practical understanding of the technology, within the hypothetical scenario of its implementation in loan applications. As DP has not been implemented in this context the game board dynamics allowed the creation of a range of DP credit imaginaries to gather consumers' attitudes towards the technology.

As participants were discussing imaginary scenarios and not their personal financial lives, as in the first study, it opened the possibility of having focus groups as opposed to one-on-one interviews [21, 90]. Focus groups allow participants with differing perspectives to discuss amongst themselves and reflecting on other participants' perspectives. As such, focus groups generate rich and detailed data [21, 90]. Had this study been done through one-on-one interviews data collection would have taken longer and the game board activity would not have worked.

The game board component was selected as an appropriate method to more easily visualise and explain the loan application process, and the interrelations between different components. This method further helped maintaining participant engagement when discussing an abstract process. Serious games, specifically role-playing games, have been used as tools to study governance in complex systems [61] due to their capacity to transfer large amounts of technical, context and processed-based knowledge, as well as helping players understand the implications of said knowledge [140],

hence being chosen as a research tool.

The DP loan game consists of a board inspired by the data flow of a loan application process (summarised in Figure 5.1), where different types of cards can be placed to create a scenario of a specific loan application situation. Changing cards and consequently scenarios ensures participants are exposed to a series of different scenarios from which the group will discuss their attitudes towards the technology.

The design of the research study builds on findings of previous PhD enquiries, summarised in Figure 7.1. In this figure the bold arrows highlight the data flow into this research study and what each of these arrows represents is described in more detail in the paragraphs below.

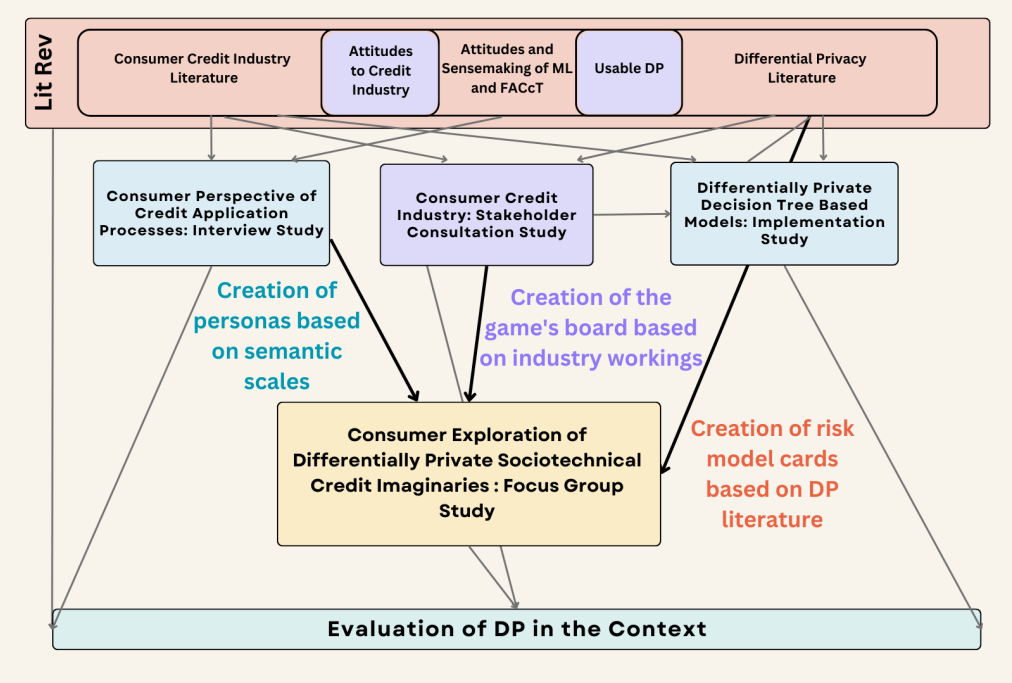


Figure 7.1: Data flow chart of PhD findings which shaped the design of Consumer Exploration Research Activity

The applicants' personas were inspired by the Consumer Interview Study (Chapter 4), specifically the previous loan experiences shared by participants. The four personas created were designed to represent consumers

from different socioeconomic, demographic and cultural backgrounds, with a variety of personas using credit to cover living costs, to others using it for their financial benefit, fitting in different market segments.

The structure of the board game was designed from the knowledge gained from the Industry Stakeholder Consultation (Chapter 5), specifically based on the data flow graph created of the application process. The findings on the preferences of the different DP behaviours explored in Chapter 5 served as a basis to design the DP models (also designed based on the literature findings which also influenced the design of the three scenarios of Chapter 5) and the game board. Furthermore, the moderator (ARP) of the game board activity used the attitudes of the industry towards DP to further question the participants by playing the role of the industry view.

The game board and interview structure were designed in an iterative process consisting of two rounds of pilot focus groups and consequent design refinements, where for each pilot there were two participants not involved in the design of the activity and two who were involved.

In order to communicate DP to the participants, the interviewer (ARP) started by explaining what a data linkage attack is. A linkage attack is an attempt to re-identify individuals in an anonymized dataset by combining that data with background information. Furthermore, the interviewer communicated perturbing data would make linkage attacks to possible to execute. A simple example and the aid of a presentation were used as communication tools. (see Appendix D.4).

Following on an explanation was given on how DP allows gathering aggregate information while maintaining individual privacy. To explain, the interviewer made use of a metaphor with some visual cues:

> DP works like looking at a group of people (representing the dataset) through some glasses (representing the model). If there is no fog in the glasses (represents noise and in this context privacy parameter) we can get the number of people in the group (aggregate information) and distinguish between different people even if they are very similar, however, if the glasses start fogging up then we can still know the number of people but can no longer differentiate between people as they start to just look like blurs the more there is fog (to represent the accuracy privacy trade-off).

An example of what a noisy answer to a query would be was also shared with the participants and a graph on the accuracy privacy trade-off, at which point participants could ask questions.

A numerical example was given (similar to the one in Figure 1.3). The query asked about someone's salary, and was based on input perturbation and hence more similar to the protections given by LDP. This type of example was chosen for the sake of simplicity and understandability for the participants and not in training perturbation (types of models that are implemented in the PhD) . The attitudes regarding DP implementation based on the different potential behaviours combined and discussed with the actual behaviours learned in Chapter 6 in Chapter 8. During game play participants simulated the randomness of DP by rolling dice, this was done to give practical knowldege on the workings of the technology. Before the begining of game play, 'participants were encouraged to ask questions throughout about DP.

The methods designed for the communication of DP made use of practical knowledge (by having the participants add randomness by rolling the

dice and observing the effects of DP) as well as metaphors, visual aids and textual explanations to explain different elements and properties of DP. Giving an overview not just on the implications of DP implementation in terms of information disclosure, but also on the randomness of the technology and the privacy-accuracy trade–off and its relationship with the privacy parameter. Karegar et al [98] created a list of functionality points, which describe base behaviours and guarantees that DP affords to be able to evaluate DP communications. While not all methods created for the focus group study address all the functionality points individually, the combination of all communications used do fulfil Karegar's functionality list. This indicates that the explanation methods designed cover all necessary information for participants to make informed decisions.

7.2.2 Materials and Procedure

This research activity was comprised of three parts:

1. **Introduction** - Participants started by introducing themselves, sharing their thoughts on credit and rating their knowledge of the industry (scale of 1 to 5 ranging from not confident to very confident on knowldege on the credit industry). This was followed by a presentation by the moderator discussing the workings of a loan application and introducing the concept of Differential Privacy. To refer to the presentation please see Appendix D.3.

2. **Game Play** - The main part of the research activity involved playing the DP loan game which involved eliciting participants' attitudes towards the different scenarios generated.

3. **Concluding Focus Group Discussion** - Interviewer led discus-

sion based on a set of summarising questions regarding the group's thoughts on DP, its behaviour as well as the workings of the Credit Industry.

In addition to the materials for the DP loan game, moderator notes, a recorder, a camera and a computer with the initial presentation were used in the research enquiry. The recorder was used to record the audio throughout the study and the camera was used to record the gameplay.

7.2.3 Game Play

Game in Brief

The game is played by rounds, where in each round there is a different model card which is randomly chosen. In each round each player rolls the corresponding model dice to get their individual private credit score. Participants discuss the impact of the different DP behaviours.

Objective

The objective of this game board activity is to communicate the workings of the loan application process and DP, as well as, gathering participants' attitudes towards different DP behaviours implemented in the credit context.

Equipment

The materials for the game activity consist of:

- **Board** - Summarises the loan application decision steps. Designed based on findings from Chapter 5. See Figure 7.2.

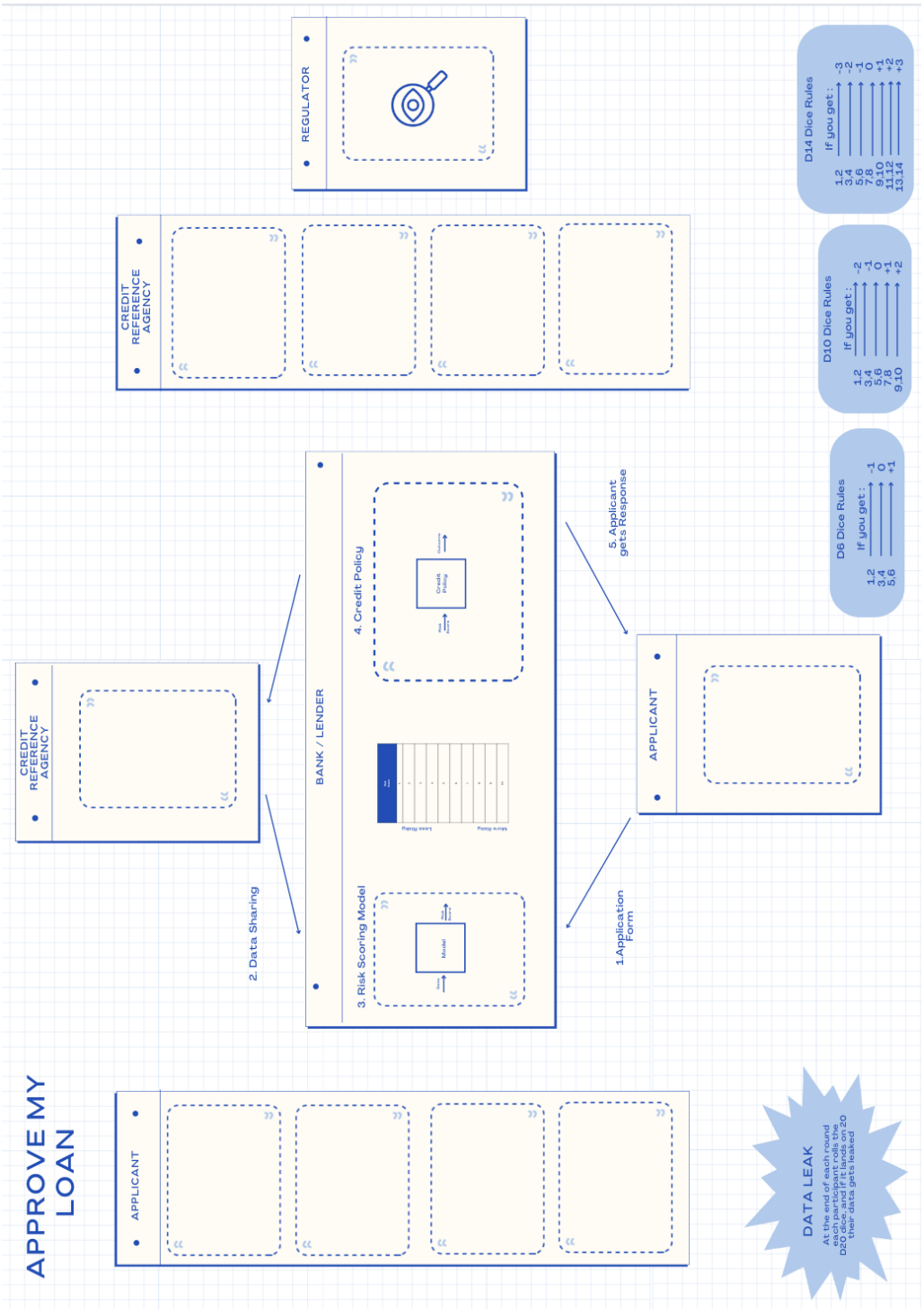


Figure 7.2: Board for the Game designed for the Focus Group

- **Die** - Different sized dice (D6, D10, D14, D20) to simulate randomness and probability. The D20 dice is used to roll for a data leak.
- **Cards** - different card types: Applicant (designed based on findings and semantic scales from Chapter 4), Data, Model (designed based on the literature of DP), Credit Policy, Regulator (designed based on Chapter 5 findings), Data Leak. See Table 7.1 for a description and Appendix D.5 for full set of cards.

Card Type	Description	Number
Applicant	Contains the name of a persona, their background and why they are applying for a loan	4
Data	Data held by CRA of each of the personas. Data fields include income, council tax, expenditure, number of credit products, number of credit applications in the last 6 months and bank account transaction data.	4
Model	Model cards describes how applicant's risk score is altered as a result of DP. Each card has a different behaviour ranging from small equal impact to significant disparate impact	5
Credit Policy	Credit policy card has a series of credit products with different characteristics, and it states which credit risk segments have access to each product	1
Regulator	Regulator cards are played by the moderator, and they request participants to discuss and justify certain aspects of the application process	3
Data Leak	Card detailing the impact of the data leak to each of the personas this card gets played if players land on 20 after rolling D20 dice after each play	1

Table 7.1: Types of Cards in Board Game

To replicate the random aspect of DP and for participants to get their post-DP risk score within the game, a set of dice is used. The 3

different sets of dice (D6, D10, D14) were used to replicate different amounts of noise added, which represented different privacy parameters. The selection of dice was dependent on the rules of the model cards, see Figure 7.3 for example.

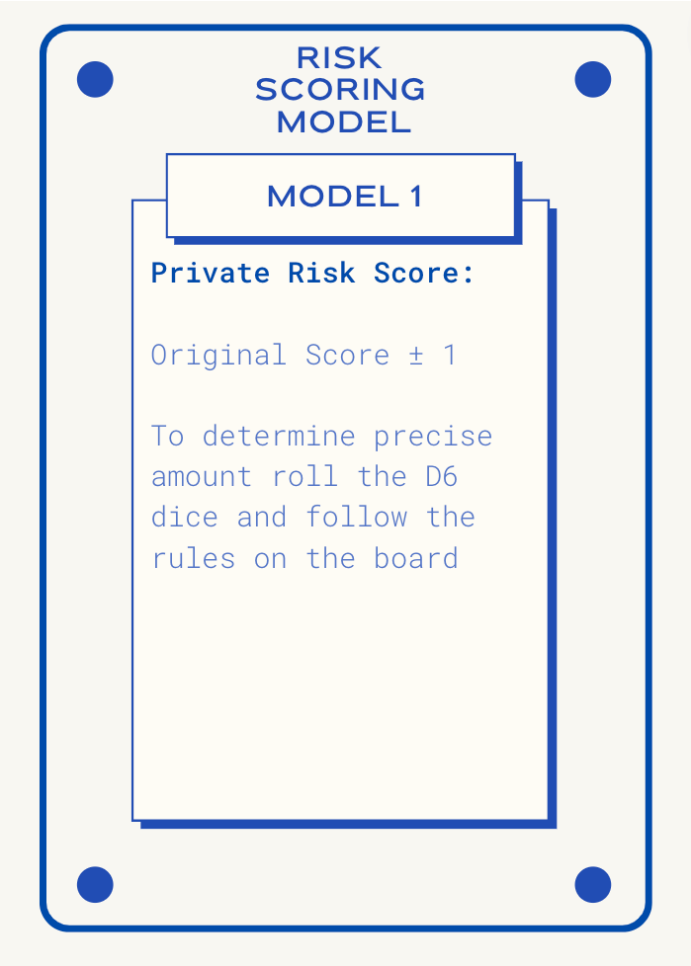


Figure 7.3: Example of a Card from the Game

The applicant cards and credit products (part of the credit policy card, each have their own APR and credit limit) design were aimed at having a diversity of applicants to represent the different market segments (subprime and prime consumers) and corresponding credit products. Table 7.2summarises the applicant ('persona' cards and the associated data cards). For the terms of each of the credit products refer to Appendix D.5

Applicants' Card	Data Card
Rosie is a 70-year-old lady who has lived in Mansfield most of her life. She worked as a nurse for 30 years and retired five years ago. Rosie never married or had children. Rosie is looking into getting a credit card to help her spread out the cost of big purchases and help with the living cost. Original Risk Score: 8	Income: 9984£/year Council Tax: 100£/month Expenditure: 752£/month Number of Credit Products:0 Number of Credit App in the last 6 months:0 Bank Account Transaction Data: Pharmacy: £24 Tesco: £13 Bus: £2.4
Mohamed (Mo) has just finished university and is waiting for his graduate program to start before it starts Mohamed has applied for a small personal loan to pay for a holiday with his university friends. Original Risk Score: 5	Income: 4000£/year Council Tax: 0£/month Expenditure: 200£/month Number of Credit Products:0 Number of Credit App in the last 6 months:0 Bank Account Transaction Data: Wetherspoon's: £14 Tesco: £37 Depop: £26
Adam is 40 years old and has been working in Data Science for the last 10 years. He has a wife and a daughter. He has a series of credit cards which he gets for the joining benefits and uses for the interest free period. He is looking to get another credit card. Original Risk Score: 2	Income: 65000£/year Council Tax: 220£/month Expenditure: 3000£/month Number of Credit Products:2 Number of Credit App in the last 6 months:0 Bank Account Transaction Data: M&S: £154 Toy Store: £56 Petrol: £62
Natasha has recently gotten divorced, and she has two children, one of them transitioning using the private health system. Natasha has applied for a loan to cover medical expenses. Original Risk Score: 6	Income: 33000£/year Council Tax: 220£/month Expenditure: 1456£/month Number of Credit Products:1 Number of Credit App in the last 6 months:0 Bank Account Transaction Data: Pharmacy: £36 Lidl: £121 Petrol: £62

Table 7.2: Applicant and Data cards

- **Pawns** - To keep track of pre and post privacy scores

Set up

Each participant is randomly allocated an applicant and a different data cad, see Table 7.2. The interviewer has all the model and regulator cards. The interviewer starts by setting the credit policy and credit cards and randomly choosing a model card.

Playing the game

The game is played in rounds:

1. Each participant reads their applicant card, and correspondent data card (only in first round) and rolls the dice to see what the final risk score is with a private risk assessment model. The participants are then asked to discuss how the change in risk score affects their applicant's lives.
2. At the end of their turn, each participant rolls the data leak dice, and if they land on 20, it starts the data leak special event.
3. The next participant at the table has their turn (direction is irrelevant as long as consistent throughout the game)
4. After all participants have discussed their applicants the group is asked about their attitudes towards the model played.
5. The interviewer plays a different model card to start a new round.

- **Data leak** Once a participant gets a 20 on the data leak dice roll, they read the data leak card which details the implication for each of the game's applicants. The interviewer then prompts participants to

reflect on measuring if DP implementation is worth protecting from a data leak.

- **Regulator cards** The interviewer can play Regulator cards at any point to start a group discussion on different topics such as: data sources used, credit policy, and the risk models. Once a regulator card is played participants are asked to put themselves in the role of the industry and justify the workings of the topic of the card (to simulate principle based regulation). Finally, participants are asked to volunteer their personal views of the topic.

After all Model and Regulator cards have been played the game ends and interviewer asks a set of summarising questions.

7.2.4 Participants

The study consisted of a total of five focus groups each with four participants held at the University. Each focus group has 4 participants to correspond with the four personas created.

The study had a total of 20 participants, with each focus group having a total of 4. There were no specific recruitment criteria apart from being fluent in English, being able to consent and being over the age of 18. All participants were recruited via a market research recruiter local to the area, who provided age and gender demographics from the participants.

Participants had a wide range of experiences and attitudes toward credit with some never having had a credit product before, participants who would rather save up or borrow money from friends over getting credit. On the other side, there were also participants making the most of different credit

products in order to get their financial benefits without paying interest. However, most participants were somewhere in the middle with either having a mortgage, a couple of credit cards and having had cars financed, with some knowledge of the process but without a lot of detail.

7.2.5 Analysis

The interview audio data was transcribed using an automated transcription service. The transcripts were manually checked by comparing them to the audio to ensure accuracy. When it was not clear which gameplay participants were discussing the video recorded during the focus group was used to clarify context.

The qualitative data elicited was analysed following the six phases of thematic analysis defined by Braun and Clarke [27] and the framework method [73], as discussed in Chapter 3.4. The data was reviewed. Initial codes were generated inductively, and once all interviews were coded; they were aggregated to form candidate themes. These were then reviewed and reorganised into their final versions. The analysis process was discussed with AL for feedback and validation.

Each focus group had a duration of around two hours, where between an hour and an hour and a half was spent on gameplay.

7.3 Results

The focus group analysis produced a series of themes:

- participant's feelings towards game applicant personas (Theme 1)

- their attitudes towards the loan application process (Theme 2)
- conflicting feelings regarding the workings of the credit industry (Theme 3)
- participants' views towards DP, its possible different behaviours, and its implementation (Theme 4)

The following section describes the data elicited from the participants and Table 7.3. summarises the main themes and core sub-themes that provide insight into the research enquiry, derived from the analysis of the focus groups' transcripts.

Table 7.3. :Theme's Table

Theme	Subtheme
T1: Participants infer and create background stories for applicants and empathise with them throughout game situations	**SbT1a:** Pts infer and create background stories for applicant's based on their financial and applicant cards
	SbT1b: General disagreement with Mo's ORS and agreement with the remaining applicant's RS
	SbT1c: Some participants have some judgement over how people spend their moneys while others mainly care if they can afford their expenditure
	SbT1d: Varied range of impact of DP on applicant's RS, access to credit and lives
	SbT1e: Data leak has financial and social implications on applicant's lives and at times their communities
T2: Agreement with loan application processes in terms of types of data used and factual decisioning, however mixed views regarding granularity of the data sources	**SbT2a:** Agreement with factualness and equality of procedures but need by some Pts of integration of some face-to-face contact
	SbT2b: Agreement with data sources used but mixed views regarding granularity of the data
T3: Tension between understanding reason for current credit policies, and their need for the existence of the credit industry and feelings of unfairness towards consumers	**SbT3a:** Credit seen as a useful tool if affordable and well managed but with possible impactful negative consequences
	SbT3b: Unfairness of risk-based pricing credit policies as those that can afford pay the least
	SbT3c: Current credit policy practises make sense from the perspective of the lenders as their main driver is profit
	SbT3d: Changes in credit industry over the years and people's relationships with cash
T4: Diverse views regarding DP implementation due to its perceived unfairness and possible impact on consumers	**SbT4a:** Range of different views regarding DP implementation, some are against it while others see it as worth it to protect from a data leak depending on RS range
	SbT4b: From the models that have exceptions rules Pts prefer bigger range to fall on higher income applicants as this will impact them less
	SbT4c: Models with bigger range have a higher risk of impacting applicants by moving them further away from products originally meant for them, however for applicants in the boundary even small ranges can impact them
	SbT4d: Randomness aspect of DP seen as unfair, and could cause applicants to behave differently

T1: Participants infer and create background stories for applicants and empathise with them throughout game situations

Throughout the game, participants empathised and put themselves in the shoes of the game applicant personas. They created possible live stories and backgrounds for the applicants based on the data given in the cards and their own assumptions and explored how the different game scenarios impacted each fictional persona.

Based on the financial cards for each of the applicants, a subgroup of participants disagreed with the Original Risk Score (ORS) assigned to Mohammed. Instead, Mohammed's ORS should have been higher considering his income, especially when compared with other applicants such as Natasha. Apart from Mo, there was a general agreement with all other applicants' ORS (SbT1b), this reflects game design which could be improved if the game were to be adapted in the future.

> *"**Pts 15:** You would think because she's [Natasha] got a well-paid job and she has more outgoing she's shown she can afford to pay back more. She should have a lower risk score because she's proven that she can afford to pay these things back, whereas Mohammed doesn't have any of that, so it should be the inverse."*

Similarly, participants made inferences and assumptions about applicant's lives based on their financial data (SbT1a), for example assuming:

- Natasha has a good house due to the high income tax
- Rosie is sick due to her pharmacy expenditure
- Mo goes to Tesco because he is a student and consequently too lazy to walk to a cheaper store

This shows participants' personal bias and assumptions based on personas descriptions and financial data impacted their views of the game's applicants.

Some participants also gave alternative interpretations of the financial data. Some participants noted the lack of context of granular financial data (discussed in more detail in T2 section) and multiple possible interpretations of said data, hence acknowledging the lack of certainty and accuracy of their assumptions.

From some of the assumptions examples previously described, it provides preliminary evidence that despite the fictional nature of the personas, there is a level of judgement regarding different applicant's live styles and expenditures by some of the participants (SbT1c) where generally Mohammed and Adam are seen in a more negative light and are "made fun of" as they spend their money in non-essential items and Rosie and Natasha are seen in a more positive light.

> *"**Pts 6:** It would be risky giving Rosie that 2 grand.*
> ***Pts 5:** She is the highest risk of the four. Because of her age and income, and that is not going up, where for the other three, they could go up.*
> ***Pts 8:** Although in a way you kind of think, it's hard not to get, I don't say my emotions are getting involved, but it's quite nice to think that a lady of 70 that has worked as a nurse 30 years, never married or had the children would be able to borrow 300 at the end of the life there in the, you know, given the times we're living in. [...] But I take the points about, you know, the risk. [...] Maybe she should have that nice sofa."*

There were also participants who cared mainly about whether applicants could afford their expenditure and credit over what each applicant spends their money on (SbT1c). These findings are similar to the findings in Chapter 4: diversity of views regarding what information should be taken into account in the ideal loan application process, with one end of the spectrum prioritising context and circumstance and the other just on financial behaviour and credit history. Based on the literature on sensemaking of algorithms [16, 24] the difference in preference could be related to the perception of the loan allocation task as either a moral decision or a mechanical decision, this topic should be explored in future work to gather more insight.

The impact of DP implementation on the personas' access to credit, over all focus group, is summarised presented in Table 7.5.

Due to the random nature of the dice roll, the game scenarios were not the same cross the different focus groups. The overall statistics of gameplay over time will be consistent, as these are caused by the interaction of: the range of score associated with each model, each applicant ORS and its interaction with the credit policy. The credit policy and applicant's ORS were designed so that some applicants would be more affected due to DP in terms of access to credit products than others similar to the findings from the industry experts in Chapter 5.

From the changes in access to credit, participants discussed how this would translate in terms of impacts on applicants' lives (SbT1d). Overall, participants consider Mohamed getting access to better credit products than he originally had, to be more impactful than losing access to some of the credit products. Framing losing access to credit as forcing Mohamed to have a cheaper/more local holiday or wait for his job to start to have one.

> *"**Pts 11:** Yeah. But, you know, if the model gives you, like, moves you down to a less risky, that could mean you're given access to more money than you might be able to repay, it's irresponsible just giving more money.*
> ***Interviewer:** Yeah. That is true.*
> ***Pts 9:** Like Mohamed having access to 2000 pounds."*

Participants agreed that the bigger the credit limit the better for Natasha´s context, due to wanting the credit for her child's transition. Furthermore, participants state that Natasha would be able to afford the better credit products with her income and expenditure. Participants recognise that Adam is barely affected by DP, and even when he loses access to credit products the terms of the ones he has access to are very beneficial (no interest and high credit limit). Some participants joke about Adam's financial position of privilege but still recognise it might be unfair/he might see it as unfair losing access to credit products due to DP.

> *"**Pts 12:** If Rosie got plus two or plus three it would have taken her out of all products.*
> ***Pts 11:** And with Adam, he would only lose one if you have plus three. But he would probably feel a bit aggrieved that he'd lost that because he kind of maybe thinks he's entitled to."*

The data leak affected applicants differently and in different aspects of life (SbT1e). Natasha not being affected financially but potentially affecting the security and privacy of her family, while Rosie had big financial impacts with her insurance premiums increasing which led participants to infer that Rosie had not disclosed all medical information with the insurance company.

Participants also assumed that Mohamed is Muslim or that his family is, which could lead to social implications for Mohamed in his community as gambling is not allowed. Mohamed's gambling might also affect him financially, by impacting his chance of getting credit. There were several discussions over the scale of Mohamed's gambling and how its impact depended on this, which once again also relates to the lack of context of some financial data.

While some participants saw the impact of the data leak on Adam's life as minimal (mainly privacy concerns) others thought his salary information could be used to scam him hence potentially financially impacting him as well. Others thought it might impact him socially or in the workplace. Overall, there was a range of views and hypotheses of how the data leak impacted each of the participants, with no consensus on who had the biggest detrimental impact, and vice versa. The impacts of the data leak on each of the applicants as discussed by the study's participants are described in Table 7.4.

Applicant	Impact of Data Leak
Rosie	**Data leak event: It was found out Rosie had a chronic illness (due to bank account transaction data)** Big financial impacts with insurance premium Assumption that Rosie did not disclose all medical data with her insurance
Mohamed	**Data leak event: Mohamed gambled a few times a week** Social implications for Mohamed if he is Muslim Financial implications as gambling might affect access to credit Level of impact dependent on level of gambling
Adam	**Data leak event: Adam's salary was made public information** Minimal impact on Adam (minority of participants) Possibility of Adam being scammed Social implications in the workplace
Natasha	**Data leak event: One of Natasha's children being transgender was made public information** No financial impact on Natasha Security and privacy impacts on Natasha's family

Table 7.4.: Data leak impacts on game personas

Table 7.5. describes how the access to credit for each applicant changed (due to DP) over all scenarios generated over the five focus groups.

Applicant	**Changes to Access to Credit**
Rosie	Mainly maintained access to the original product she had Lost access to all credit a few times Gained access to a better product a couple of times
Mohamed	Lost access to original products most times Maintained access to original products a few times Gained access to a better product a few times
Adam	Maintained access to original credit products most times Lost access to a credit product a couple of times (Never gained access to better products as he was originally allocated the best credit product)
Natasha	Lost access to original products most times Gained access to a better product a few times maintained access to original credit products once

Table 7.5.: Changes to access to credit on game personas

In summary, participants engaged with the game's personas, bringing their own assumptions and biases of each of them based on the information provided. Participants were able to empathise and put themselves in the persona's situation, having complex feelings and attitudes towards the game scenarios and their impact of DP on the different personas. This in turn allowed participants to extrapolate the game scenarios to the real world and develop personal attitudes towards DP, which will be explored in more detail in T4.

The design of the game with its specific elements was effective in promoting discussions on the topic of how individuals might be affected by DP, indicating an appropriate choice of methodology [140] and successful deployment.

T2: Agreement with loan application processes in terms of types of data used and objective decisioning, however mixed views regarding granularity of the data sources

Across participants there is agreement with the current working of the loan application process.

The majority of participants agree with the loan application process being based on facts, in contrast with the added randomness of a differentially private model (see T4), which showcases a requirement and need for "objectivity" (SbT2a).

> *"**Pts 9:** That's what I meant by the I'd rather it was based on facts. Because then you're properly assessing whether you can afford to do it [get a credit card] and whether it works for them [banks] as well. Whereas DP it's sort of putting everyone slightly in the dark."*

Some participants describe a preference for the same process being used for all applicants, i.e. procedural fairness (based on Chapter 4´s family of fairness definitions). However, not all are fully satisfied with the current process and some disclose their desire for the integration of face-to-face interactions (SbT2a). These findings add further evidence to the findings from the Attitudes and Experiences with Loan Applications Study described in Chapter 4.

Over the five focus group workshops there were a range of different discussions and opinions on the data sources that are currently used and which could be used in the application process. These discussions were prompted by the interviewer playing the associated regulator card. There was a majority agreement with the data sources that are currently used (mainly income and expenditure), however not all applicants were comfortable with the level of granularity of the Bank Account Transaction data.

> *"**Pts 16:** I don't think I like them [banks] looking into what I'm buying or where I'm buying things from. I mean that Marks and Spencers it's not necessarily food. It could furniture, it could be clothes.*
>
> ***Pts 15:** Whereas for me, I would, I would expect if I was applying for a loan, I would expect the Bank to look up where my money is going. So that they could, well, hang on a second. You wanna loan for, say, 5000 pounds you're spending extortionate amounts of money at a toy store. It paints a picture of, when put together with other things like say, for example, if they're seeing lots of transactions with a betting company, [...] they [banks] can use it to help their risk assessment of you as a person.*
>
> ***Pts 16:** I don't like the idea of Big Brother, no.*
>
> ***Pts 14:** I think I'm just accepted it now.*
>
> ***Pts 13:** There's a kind of level of acceptance, I think. [...] I'd be happy to say, look, this is my income. If you want to see my accounts or my tax return fine, which is what it used to be back in the day wasn't it? But this [bank account transaction data] starts to paint pictures rather than numbers."*

The extract showcases the acknowledgement by a subset of participants on

the limitations of the data sources, as bank account data still lacks context and hence can have multiple interpretations. The extract also refers to the impact of the granularity of bank account transaction data.

There were also discussions over if lenders should take into consideration factors such as what the intended use of credit is and if people's expenditure is on frivolous items over essentials, however, there were varied views across the participants and no consensus. The arguments against this relate back to the idea of objectivity and using these data sources is perceived as subjective and could lead to biases. Furthermore, some of this data is hard to monitor in practice. Those in favour mentioned that context could help in more "fitted"/personalised loan allocation, i.e. making sure applicants are not given more money than they need and that they have the most relevant products, referring to the example of Mohamed and a travel credit card.

A few of the participants expressed both views, this is, taking intended use into consideration would be good, especially to know if for essential purchases/expenses or not but acknowledging that this would be a subjective field hard to monitor in practice. This indicates that these participants might see loan allocation as neither an entirely moral nor mechanic decision, and the need to create a process which might be able to satisfy both these aspects of objectivity and equality of procedures as well as empathy and moral decision-making. While this is a complicated task, future work could explore participatory design of the loan allocation process (and/or wider financial system processes) with the general public, as well as adaptaion of the game for financial literacy aims.

T3: Tension between understanding reason for current credit policies, and their need for the existence of the credit industry

and feelings of unfairness towards consumers

There were some participants who preferred not to use credit due to its possible negative impact, I.e. getting a loan which is not affordable can lead people into a debt spiral which impacts their credit score. Most participants see credit as a useful tool to help spread out costs over time. Applicants are able to afford long-term but they do not have the expendable income to afford as one time payments (SbT3a). In order to make the most out of credit products, consumers need to be able to manage their finances.

> *"**Pts 16:** I think it's [credit] good for people that are good with money and that need the credit. So, I'm not totally against them, but not for me."*

A couple of participants also wished access to credit to be harder in order to prevent people from getting into financial difficulties, or getting credit if they did not need it. While these attitudes are based on concern over the impacts of debt on people's lives, there is a counter argument that it also disregards people's personal agency and infantilises consumers.

While participants generally agreed with the application process, the same cannot be said about the credit policies from mainstream lenders. Risk-based pricing was generally considered unfair for the consumers as those who could afford the most paid the least interest-wise and had access to better credit products. While those who could afford the least have high-interest rates, making it harder to repay said loans and potentially getting themselves into a debt spiral (SbT3b). Concurrently to these feelings of unfairness towards the consumer, some participants also understand the implementation of the current credit policies from the perspective of the lender, as these are mainly moved by profit and need to be profitable to

continue to exist (Sbt3c).

> *"**Pts 18:** I mean these guys [lenders], I mean they're in it for the money. And if they don't make money, they're not going to be there. So, you need them to feed it in in, in the way that we live, in today's society, we need them to be there. But on the other hand, it needs to be some sort of regulation that they are not making so much money."*

A couple of participants discussed the tension being a result of the current capitalist economic system and its dynamics, and a minority of participants also suggested interest rates to be capped so that even high-risk people can access credit at a cheaper rate. Similarly to the views towards taking intended use for credit into consideration in the application process, the findings presented in this theme highlight the need for re-design and restructuring of both processes and the wider industry to make credit work better for consumers.

Over the course of the workshops participants also shared and reflected on their experiences with the financial system, and the changes that have happened over the years. Previously credit used to be considerably harder to access, and there were a lot fewer credit products on offer. Some participants discussed being able to access their credit scores via the Credit Referencing Agencies' websites and know which credit products they would be approved for, which was surprising for participants who had not interacted with the industry in recent times. These discussions highlighted the increased complexity of the industry nowadays. Some of the older participants had not kept up with its evolution due to their lack of involvement with industry. Younger participants who applied for credit products were aware to a bigger extent of the details of the workings of the industry.

Participants shared their attitudes towards the upcoming Buy Now Pay Later Credit specifically their similarly with payday loans with the high interest rates and easy approval and consequent perceived risk to consumers. Some participants noted that attitudes and knowledge towards credit might also have some generational components. Older applicants discussed the financial views passed on to them by their family, which highlighted not to buy things on credit.

Furthermore, there was also a discussion regarding the use of physical money over cards, how it has a different psychological effect as people are more aware of their spending.

T4: Diverse views regarding DP implementation due to its perceived unfairness and possible impact on consumers

Participants have varied views regarding the implementation of DP. Some of the participants do not consider DP implementation worthwhile when weighed against protection from a data leak arguing that the financial impact of always having the risk score affected is bigger compared to the impact of a potential data leak.

> *"**Pts 1:** I don't think my data is that interesting. So, the risk of me losing out financially because you know, I'm trying to be more private and then they [lenders] say no, we're going to have to charge you more interest in an effort to disguise my personal data. I would go down a couple of brackets or even one or one bracket that might have a real impact in my life. Someone knowing that I that I eat my dairy milks is only going to do a limit limited harm to me, and I think it's true for most people. Most of us live quite small lives as it were that there's*

only so much harm, they can do to us with financial data."

Other participants thought it might be worth it depending on the Risk Score Range, where the smaller the range of possible change the better. Finally, a very small minority, a couple of participants, thought any range would be worth protecting from a data leak, where even if all ranges would be worth the protection the smaller the range the better in terms of applicants' impact (SbT4a).

Across the different Risk Assessment Model cards from the game, there were three which did not have the same risk score range for all applicants. From these three models' the majority of participants prefer the possible wider change to fall on applicants with low-risk scores. As these are usually associated with higher incomes and even with a big range of possible changes, usually this subgroup of applicants would be less impacted. I.e., still has access to good credit products regardless while for high-risk applicants the difference between the credit products obtained with a high range is much more significant (SbT4b), see Adam and Rosie game scenarios in Table 7.4.

Similarly to SbT4a, participants generally prefer models that have a smaller range of Risk Scores as a small change in risk score will leave participants accessing credit products that are more similar or the same as they had before DP. However, several participants also noticed from game scenarios, see Table 7.5., that applicants with a Risk Score which is on the boundary between access to different credit products are usually still impacted even for small ranges. For example, Mohamed and Natasha had more changes to access to credit products when compared to Rosie and Adam (also in SbT1d).

There were mixed views regarding preferences over models with an equal range for everyone and differential models. Some preferred everyone to be on the same range as it was perceived as fairer due to the equality of process. A minority group preferred small range for high-risk applicants even if that meant a higher one for low-risk applicants as that was perceived as fairer due to equality of impact.

As participants value the factualness of the loan application process (SbT2a), the randomness inherently added by DP is perceived negatively. DP makes it possible for applicants to gain access to credit products that might not be suitable for them and potentially unaffordable which could cause a negative impact if defaulted on, or alternatively deny the opportunity of certain credit products to applicants that could afford them. Some participants further described the idea of earning access to credit/deserving access to certain credit products as a result of their financial actions which the randomness takes away from:

> *"**Pts 2:** I feel like for the everyday, normal, standard person, there seems to be higher risk of losing out [by implementing DP]. So as the standard everyday normal person and I don't think I would want that risk attached because these loans essentially allow you to progress in some way, right. And I feel like I would be getting an unfair judgment and risk attached to me and that isn't necessarily true to what I'm able to do in my actual life. And for me, it was kind of simply in negative experience."*

Both the ideas of earning access to credit and credit being a tool for social mobility, still reflect the same ideas put forward by advertisements and the notions of self-reliance characteristic of the Thatcher era discussed in Chapter 2.1.

The awareness of the implementation of DP and specifically the randomness aspect might make consumers interact differently with the industry, by for example applying more often when denied loans as the randomness might lead to them being allocated better credit products.

> *"**Pts 10:** If it's random, every time you applied, you know. So, like if you applied once, you might not get something, but if you applied again and again even though your circumstances haven't changed suddenly, maybe you get something you couldn't get before.*
> ***Pts 11:** That might encourage the way to think. I'm gonna just keep applying every six months because I know that there's this thing going in the background."*

Furthermore, participants thought that the general public would generally not be very happy with the implementation of DP. The idea of DP being optional was discussed by a couple of groups but while participants liked the idea of choice, they were worried about how the banks would treat consumers who chose DP, as they could potentially correlate with risky/illegal behaviour as described by participants.

Overall, the implementation of DP was conditional on its behaviour as the negative impacts, changes to access to credit, were considerable when compared to the privacy protection afforded by the technology. Hence only for small ranges of variance to the risk score would DP be worth implementing for some but not all participants.

7.4 Discussion

This study gathered users' attitudes towards the implementation of different DP Risk assessment models in the context of credit loan applications. In order to achieve this a game board activity was designed to create a variety of different tangible scenarios to analyse the impact of the different models on the individual personas and group effects. Table 7.3. summarises the findings of the research enquiry.

The existing literature on Usable DP (see Chapter 3.2.5.) gives insight into testing different ways to communicate DP and their impact on users' willingness to share information [175, 103, 46]. In the context of this thesis, the question of impact of DP on willingness to share personal data is not relevant, as that is something applicants do not have agency over.

While DP has been studied in specific industrial contexts such as healthcare [8] this has not gathered user input either. As highlighted currently there is no specific literature to be able to compare this study's results to.

This work contributes to the understanding of consumer's attitudes towards DP and how to communicate it. This has been done through the design of novel communication methods, including the visual glasses metaphor used in the initial presentation (Appendix D.3) and the interactive DP loan game. The latter further communicates the complex process of loan allocations to naive individuals. This is the first study that gathers users' attitudes towards different models' accuracy behaviours.

The aim of this research enquiry was to develop new knowledge on consumers' attitudes towards DP implementation in the loan application process. Overall participants had a variety of views regarding the implementation of DP, which were dependent on the range of variation of the risk

scores which were a proxy for different levels of accuracy drop. There was a preference for smaller ranges due to less impact on applicant's lives. From the differential models, a preference was communicated for the bigger range to fall on applicants with a lower risk score. Finally, the random aspect inherent to DP was seen as unfair to the applicants as it adds a degree of uncertainty to a process that, for the participants of this study, should be based on facts and the same for every applicant.

The study found mixed views regarding the implementation of DP, as it is not seen as that beneficial due to the impact of the accuracy drop on applicant's lives.

As participants' sensitive data is already being used in the loan application process, one can speculate that there is a base level of acceptance of said data use. This could explain the perceptions of the implementation of DP in the loan application process as not that beneficial.

The pros of DP, specifically preventing/making less impactful a data leak and future possible inference attacks, are not seen as worthy enough against the negatives.

As a result of implementing DP in the Risk Assessment model one could in theory perform third-party querying to explain individual outcomes of the loan application, a need expressed by participants of Chapter 4. The DP implementation could prevent the risk of reverse engineering the model and consequently inadvertently sharing the lenders' intellectual property. This application of DP should be investigated in future work.

Similarly to the findings in Chapter 4, there is a general agreement with the loan application processes where some smaller groups of participants want more face-to-face interaction. Furthermore, the varied views regard-

ing the granularity of data sources used in the process supports the findings of Chapter 4 where participants of that study also mentioned their desire for explanations on how different data sources are used. While there are similar points in the two studies the interview study (Chapter 4) has more heterogeneous views while this focus group study only has a smaller subgroup of participants that share the same views. This could be a result of group dynamics and an inherent limitation of the focus group method.

There were also some feelings of tension regarding the current credit policies used in the mainstream financial industry, with an understanding of the perspective of the lender as a business focused on profit, but concern over the impact on consumers, especially due to risk-based pricing. The results support, specifically Theme 3, the regulatory implementation of caps to certain credit products in order to protect the consumer, as well as a possible extension of these caps for high-interest rate credit products. This evidences the need to create a process which might be able to satisfy both lenders and consumers and the diverse needs for both objectivity and empathy expressed in Theme 2 and Theme 3 as well as in findings from Chapter 4.

7.4.1 Reflection on Study Design, Limitations and Future Work

For this research enquiry, a focus group activity based on an interactive serious game was chosen as the most appropriate methodology. The game component was chosen due to its potential to communicate complex processes and knowledge [61, 140], such as the implementation of DP in a loan application scenario. A range of credit application scenarios were generated through gameplay, and became the basis for the enquiry into participants'

attitudes towards DP. Overall the choice of interactive activity was successful, as evidenced by participants' empathy and interaction with personas, highlighted in T1. Participants were able to speculate on the different models' impact on the different personas, showcasing an understanding of and consequences of DP. Furthermore, applicants were able to extrapolate the process-based knowledge gained from the activity to form their own personal attitudes towards the technology.

Some of the specific elements of the DP loan game could however be further re-designed and improved: the original risk score attributed to Mohamed was perceived as not being reflective of the financial and applicant data provided..The data leak event could be more reflective of the privacy guarantees that would be afforded if DP were implemented within the risk assessment model, for example the private explanation of outcomes described in the previous subsection.

The DP loan game could be extended to have a wider pool of applicant personas and could be used as a research tool in co-creation activities of DP, by having participants design their own model cards and test them using the game, or alternatively designing a different loan allocation process by redesigning the board itself.

The data from this study was collected from a total of five focus groups each with four participants. Focus group studies tend to have a larger number of groups [30], in future extensions of this study, an additional three to five focus groups would be suggested, dependent on results saturation. With the five focus groups of this study, the results were starting to show clear trends and repetitions, but the suggested additional focus groups would serve to confirm the saturation of data. The data collected from this research activity might not represent participants' views fully, as they might

feel less inclined to share some of their thoughts especially if these differ from the group consensus or if they are less generally socially acceptable [30, 44].

All participants were recruited using a market research recruiter who only collected age and gender demographics, hence the study's population might not be representative of the UK population. Demographic data was not collected as it was not planned to be used in the analysis, but further iterations of this study should be in order to assess representativeness of the UK population.

Figure 7.4 summarises the reflections on what went well and should be repeated and what could be improved and expanded in future work.

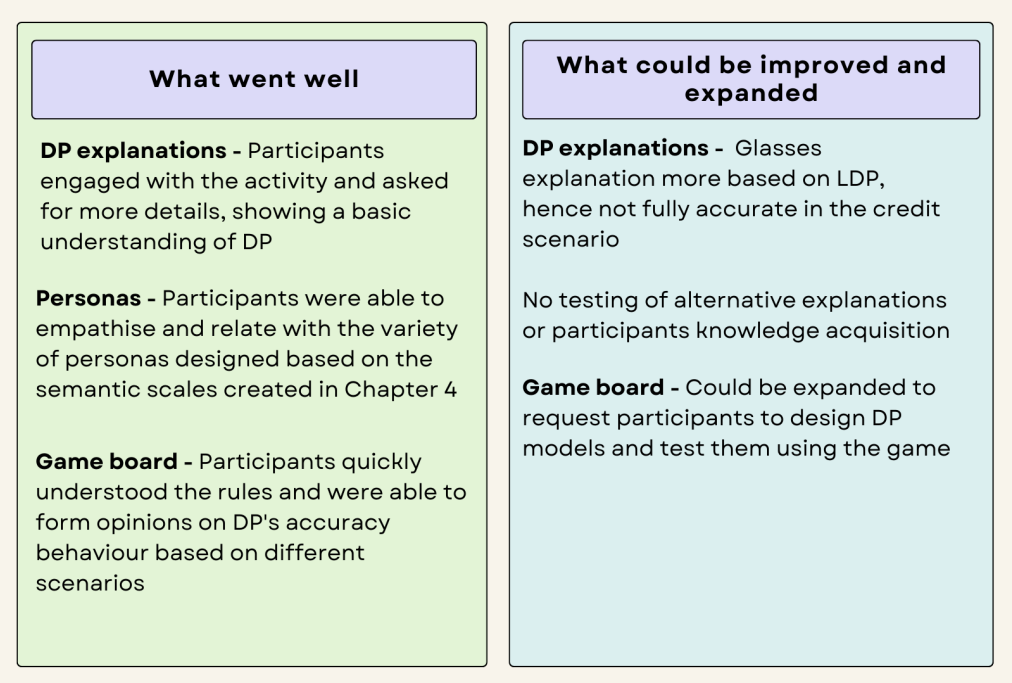


Figure 7.4: Consumer Exploration of Differentially Private Credit Imaginaries Reflection

This work has contributed to the DP communication field with the design of a novel DP communication with the glasses metaphor, making use of visual explanations which are not overly abstract. Participants seemed to respond

well and understand the basic principles of the technology based on this explanation, their understanding was then further consolidated by playing the DP loan game. However, no specific measurements of knowledge acquisition and increase in user understanding were carried out. In future iterations of this research study, alternative DP communications should be considered in the pilot stages and measures of increased understanding and knowledge acquisition should be carried out.

As this is one of the first studies of consumer's attitudes towards accuracy behaviours in DP implementations this work should be extended in different contexts, in order to better understand users' requirements and needs regarding DP.

7.5 Summary of Findings

Overall participants had a variety of views regarding the implementation of DP, which were dependent on the range of variation of the risk scores which were a proxy for different levels of accuracy drop. There was a preference for smaller ranges due to less impact on applicant's lives. From the differential models, a preference was communicated for the bigger range to fall on applicants with a lower risk score. Finally, the random aspect inherent to DP was seen as unfair to the applicants as it adds a degree of uncertainty to a process that, for the participants of this study, should be based on facts and the same for every applicant.

In the following Chapter, the results of this study will be combined with the others and discussed in detail to reach the final evaluation and thoughts on the implementation of DP in the Risk Assessment model of a loan application.

Chapter 8

Discussion

8.1 Introduction

In this chapter, I will summarise the outputs of the PhD, articulating them in relation to the original research question, new knowledge and wider findings that cut across the different research activities. I will critically review and discuss limitations of the thesis´ work. Finally, I will discuss in depth its contributions and possible future work.

8.2 Summary of Research

As presented in Chapter 1 this thesis' main aim was to explore the possible implications of implementing DP in the Risk Assessment Models of consumer credit loan applications from a user perspective. This led to the development of the overall research question: ***What are the repercussions to costumers of the implementation of DP in Credit Risk Assessment Models in UK consumer credit industry applica-***

tions?

Due to my philosophical positioning between positivist and social constructivist, and consequent acknowledgement of the mutual constitution of society and technology, an approach to research based on the sociotechnical premise was taken, as explained in Chapter 3. This informs the PhD as it accounts for the complexity and nuance of technological implementation in an applied industrial setting.

Based on this premise five sub-research questions and associated research activities were defined and designed as the building blocks to achieve the overall aim. This thesis takes a mixed-methods approach, with a larger contribution from the qualitative perspective which engages with a variety of relevant stakeholders. The relevant literature for this thesis is summarised in Chapter 2.

Chapter 4 describes the Attitudes and Experiences with Loan Applications study, which was designed to answer ***What are consumers' attitudes regarding current loan application practices?*** via a semi-structured interview and a follow-up survey of people with experience applying for loans in the UK. Chapter 5 discusses the UK Consumer Credit Industry Stakeholder Consultation study, where a variety of stakeholders from the industry were interviewed to gather data to answer ***What are the processes of the consumer loan application that impact outcomes?*** and ***What are the UK consumer credit industry perspectives on DP implementation in the risk assessment model of the loan application?***

What is the accuracy drop behaviour for DP Decision Tree based models applied to credit risk assessment models? is answered through the implementation of a variety of differentially private algorithms

described in Chapter 6. Finally, Chapter 7 reports on the group game-based interactive research activity. This activity was designed to expose consumers to a variety of potential sociotechnical scenarios, created based on the literature and findings from the three previous research activities. The activity answers ***Exploration of consumers' attitudes towards DP implementation in the loan application process based on an interactive game board.*** research question.

These research studies have culminated in a new depth of understanding of the social perspectives and potential technical impacts of an hypothetical DP implementation in the risk assessment model of the credit loan application process and associated decision-making.

8.3 Reflections of overall approach and research limitations

Reflecting upon my PhD journey and experience, the change at the start from a positivistic approach to the study of DP to a more user-centred and constructivist approach necessary to be able to address the challenges of this technology in a practical setting. It's necessity became clear in my mind as a result of the literature review, specifically the disparate accuracy drop in the DP-SGD models presented in Bagdasaryan et al [15], Jaiswal et al [95], Xu et al [176] and Farramd et al [64]. These prior works showcased how the privacy-accuracy trade-off highly depended on the combination of algorithm and dataset and could potentially impact subgroups of users very differently.

However, this change in approach also lead to its challenges namely the

need to learn and master qualitative methodologies. There was a the lack of similar approaches to the study of DP, at the time, to help with this learning process.

Learning about qualitative methodologies as I was designing and executing the initial research studies has had an impact on the methods used, i.e. both the industry consultation and the consumer initial exploratory study were interview based. Interviews are a standard approach to qualitative methods, hence my initial choice of this method took a more conservative approach. Furthermore, the initial analysis of the Consumer interview study was reviewed and rebuilt to make sure it had enough depth and attention. On reflection, I could potentially have done a quantitative follow up survey to validate the findings summarised in the semantic scales. This quantitative validation of the semantic scales would provide important insights to stakeholders in the credit industry as well and may have led to further application of this framework.

My confidence and skill with qualitative methodologies based on the learning process throughout the PhD is highlighted in development and undertaking of the Focus Group activity. Incorporating an interactive game board activity is not a standard approach and is especially novel in the ML/DP setting. This demonstrates my understanding of the issues a focus group investigating DP might encounter as well as knowledge about potential tools and stimulus materials to help address these challenges. In the Future Work section of this chapter I will delve into the potential of adapting the game for educational purposes.

Due to the highly interdisciplinary nature of my approach, the work often struggles to find an already established academic home. While there is evidence (see Chapter 2.2.5.) that my work fits the area of Usable DP

and associated interdisciplinary research, at the start of the PhD this area was not as developed. Therefore, there were no specific methodologies and approaches to follow. This added to the challenge of understanding how to account for the sociotechnical system in the research design and which qualitative methodologies could help achieve this. Instead, for the development of the studies presented in this thesis I took a more pragmatic approach by searching various academic areas and industry contexts (such as HCI, HF, ML and so on) for approaches and methodologies that might fit the research aim.

Overall the inquiry presented in this thesis is an unusual approach to the study of a technology not yet implemented. It combines highly technical elements, specifically the DP Decision Tree implementation study, with qualitative work on the processes of the industry itself and about consumer attitudes towards it. Approaches like this and sociotechnical evaluations tend to be done by interdisciplinary teams, as discussed in [43]. This was not the case due to the PhD context and led me to undertake the challenge of learning a new set of skills and approaches to be able to address the challenge of studying technology from a sociotechnical perspective. Usually within an interdisciplinary team each member might have a different perspective and expertise leading to the overall evaluation being more complex and varied. Due to my academic background several possible elements were not investigated in depth and could be pursued in further work such as the economic side of the loan application process and of the industry dynamics.

Furthermore, no other PET were considered for the study as DP was already established as the state-of-the-art method, falling in the reductive view of privacy highlighted by Sarathy [144].

Had I taken a more positivistic approach this thesis would have focused more on the technology itself and potential ways to decrease the privacy accuracy trade-off, and it would probably have more quantitative studies over semi-structured interviews. The more constructivist approach based on understanding and empathy led to a user-focused approach, with enquiries into the experiences and needs of the consumers, as well as, their attitudes towards the technology in question.

Throughout the study chapters' Limitation sections, I explore the specific limitation of each study in more depth and provide opportunities to overcome these limitations . Some of the main limitations across studies involve:

- participants potentially tailoring their responses based on their expectation of what the researcher (ARP) wants to hear (confirmation bias)
- the small sample size as a result of interview methods and the inherent closed off nature of the industry which stifles research
- the inherent and unconscious impact of my lived experience and perspective on the qualitative analysis and moderation (researcher bias)

Some of these limitations such as small sample size are inherent to the methods chosen and limit the generalisation of the findings [30]. This could be overcome by following up the interviews with a quantitative survey. The remaining limitations stated above can be addressed by a reflexive and critical approach to the interviewing. In addition, paying attention to the language used to avoid influencing participant's answers.

Furthermore, the studies presented in this PhD were conducted during the Covid-19 pandemic. Different studies were conducted at different times of

the pandemic and hence under different lockdown restrictions. Study 1 and 2 were both conducted online, as at the time, this was the only viable way.

During this time a variety of research activities has had to be alted or adapted to the social distancing restrictions [100, 133, 110]. For study 1 and 2 this meant an exclusion of observational methods. Performing the studies online might have affected the ability to quickly create a rapport with each participant. Archibald [7] reflections on online interviewing highlighted the occurence of technical difficulties. Access to a stable connection, which might be a proxie for socioeconomic status could have influenced the quality of the interview and consequent analysis.

Furthermore, as highlighted in section 2.1.3. the Covid-19 pandemic and associated lockdowns had big financial impacts on the general public [12]. This in turn might have affected the consumers' experiences and perceptions of the credit industry.

Despite the some of the limitations of the work presented and discussed here, this thesis' findings bring some important contributions that will be discussed in the next sections.

8.4 Summary of Findings

In this section I will summarise the main findings, by answering the sub-research questions as well as outlining other findings that cut across the different research activities.

8.4.1 RQ1: What are consumers' attitudes regarding current loan application practices?

From the first research study (Chapter 4.3.1.) we learned that the application process is user-friendly, however applying for loans can be an emotionally charged experience independently of the ease of process due to its potential impact on applicants' lives (Chapter 4.3.2.).

Regarding the use of traditional and novel personal datasets in the decision-making process for loan application outcomes, there is conditional acceptance if the sources are perceived as useful for the application from participants of research study 1, the user interview study, (Chapter 4.3.1.). This was additionally supported in views discussed by participants of research study 4, the interactive focus group (Chapter 7.3.). Participants from both studies have some feelings of discomfort with the granularity of the detail captured by some of these sources, leading to feelings of tension.

This means that in order to maintain and improve the industry's efforts in improving consumer trust [4] they might need to develop more effective communication on why each data source is collected, with whom is that data shared and which measures are in place for it to be protected. Alternatively, the industry might need to reconsider, based on consumer consultation, which data sources should be included in the process and which do not add significant benefits and can be removed.

Xiong et al [175] and Kuhtreiber et al [103] found evidence that the implementation of DP leads to consumers sharing more of their personal data ,as such, the implementation of this technology within the loan application process might appease the discomfort with the amount of personal data required.

Automation applied to loan applications elicited diverse views on the following topics: what fairness is, modes of interaction, and what should be taken into consideration as part of the decision process. This diversity is consistent with the literature [148, 16, 106] and showcases different underlying values prioritisation (Chapter 4.3.3.). I.e. some users value accuracy and objectivity over empathy while others the other way around. As discussed in Chapter 4 where one falls on this spectrum might be related to perceiving loan allocation as a moral or mechanical decision, adding further insight to the findings of Bigman and Gray [24].

The views of participants from the game board focus group activity (Chapter 7) are less heterogeneous. There is a general agreement with factualness and equality of procedures (with a subgroup wanting more face-to-face interaction) and feelings of tensions regarding the credit policy, i.e. they make sense from the perspective of the lender but feel unfair towards the consumer.

Finally, consumers have a good understanding at the macro level but would like more transparency regarding micro-level details (Chapter 4.3.1.).

Addressing the variety of participant views on how decisions should be made might only be possible by changes at the wider industry level, such as with an increase in credit unions and sustainable lending companies. At an industry level, users who see loan allocations as mechanical tasks will be drawn towards traditional lenders as these already make decisions based on factualness and equality of procedures based on people's credit scores. Those who perceived loan allocation as a moral decision might be more drawn towards credit unions and sustainable lending companies, which can take into consideration people's backgrounds. While some of these companies already exist, more funding and initiatives by the regu-

lator and government is essential for them to become a viable alternative [172]. Furthermore, advertisements and educational campaigns about these alternative sources of credit would be beneficial so more consumers see them as options to consider. These findings provide further evidence that adds value to the call for increased presence of alternative and responsible lenders in Woolard [172].

An initial approach to try and address the variety of user needs in terms of both objectivity and empathy at an individual lender level, could explore participatory design of the loan allocation process (and/or wider financial system processes) with the general public, as discussed in the Future work section of Chapter 7.

8.4.2 RQ2: What are the processes of the consumer loan application that impact outcomes?

Currently, credit is a competitive expansive market where profit is the main driver but balanced by a principle-based approach to regulation focused on consumer impact (Chapter 5.3. T1).

The loan application process is made up of a credit risk component (built based on conversations around the right balance of complexity and accuracy) and credit policy (based on NPV models which gives the value of a loan to lenders over the course of its lifetime) and involves several stakeholders (Chapter 5.3. T2).

Within the industry, technology such as ML and AI is seen as useful but is implemented conservatively to maintain the right balance between complexity and accuracy (Chapter 5.3. T3). Under the current working of the industry accuracy and profit are linked and related through and due

to the existing credit policies (hence being profit that drives technological implementation) (Chapter 5.4.). These findings add value the literature [172, 130] as they provide richer insights due to the method chosen, interviews.

Overall, all stakeholders (industry, regulators and third sector organisations) agree on the importance for consumers to be able to access credit but have differing views on how this should be achieved. This creates a dynamic balance between profit and expansion (lender priority) and consumer impact and protection (regulator priority) (Chapter 5.3. T1).

The data captured from the stakeholder interview study in response to this sub-research question supports findings on the industry stance on technology [13] and adds a more in-depth understanding of the underlying causes and motivations for it. The findings from this research question are elemental to the understanding of the sociotechnical system of the consumer credit industry, by relating organisational structure and processes to decisioning on technology implementation. While we apply these findings to the evaluation of DP later in this chapter, the knowledge provided is relevant to all those interested in credit risk modelling technology and the wider financial sector.

8.4.3 RQ3: What are the UK consumer credit industry perspectives on DP implementation in the risk assessment model of the loan application?

There is a general agreement amongst the industry participants that a disparate accuracy drop of different consumer subgroups would not impact them equally. However, different participants had different views over

preferred behaviours. There was also agreement that subgroups in the extremes of the risk score range would be less impacted overall (Chapter 5.3. T4).

Regarding the management of the accuracy-privacy trade-off, some participants think the privacy parameter should be set by the regulator, others by the lenders (Chapter 5.3.). This indicates that previous to the implementation of DP in the context, or as a result of its wide in the industry, there would potentially need to be a consultation between industry and regulators to better understand how to manage the inherent accuracy-privacy trade-off.

The implementation of DP would be dependent on the amount of accuracy loss and regulatory encouragement. At this time, the regulator is unlikely to give said encouragement as that would differ from their principle-based approach to regulation and due to the potential negative impact on consumers due to the randomness aspect of DP.

8.4.4 RQ4: What is the accuracy drop behaviour for DP Decision Tree based models applied to credit risk assessment models?

Generally, models handle numerical variables similarly well, however not all perform well with categorical variables.

Models should be chosen based on good accuracy and steep climbs of the privacy accuracy graph, to be able to set a relatively low privacy budget and maintain accuracy (Chapter 6.5.).

DPGBDT 1x100 is the best-performing model across the different datasets

where, for $\varepsilon > 4$, it would approximately maintain the accuracy compared to the non-private model implemented. For the Heloc, which is the most realistic dataset, for large privacy budgets, this model surpasses the non-private algorithm's (LightGBM) accuracy (Chapter 6.4.1.).

DAL of specific subgroups across all datasets and models implemented are minimal, however, ODAL is more common (Chapter 6.4.2.).

Based on these findings, were DPGBDT 1x100 implemented it probably would not lead to a massive decrease in accuracy as the datasets used in practice are mainly numerical. In the case of lenders which might still base their risk scoring models on more classical statistics such as LR, the implementation of DPGBDT 1x100 might actually result in an increase in accuracy compared to their previous model. Employing novel ML explainability methods in hand with DP might help mitigate the increased uncertainty in the model due to the noise added by DP.

8.4.5 RQ5: What are consumers' attitudes towards DP implementation in the loan application process?

Overall participants had varied views regarding if DP should be implemented, as privacy protection is seen as important but might not compensate for the potential negative impact of DP on credit decision outcomes (Chapter 7.3. T4).

The randomness element of DP was seen as unfair, as participants prioritise factualness and equality of process. According to Balies et al [16] algorithms are perceived as fairer than human agents when performing mechanical

tasks due to their objectivity, hence the implementation of DP and its associated randomness seems to deter the benefit of algorithms for loan allocation.

Across the different DP accuracy behaviours discussed, models with a bigger range are seen as riskier as applicants would be given products that are very different to the originally allocated ones. If a model does not have the same accuracy drop for all subgroups and consequently different ranges, participants would prefer for the bigger range to fall on applicants with higher incomes as they will be less impacted "in practice" (Chapter 7.3. Theme 4), this indicates people don't want credit decisions to perpetuate inequality.

As an applicant's risk score affects the types of credit products they have access to (risk-based pricing), the change to the risk score financially impacts participants and is seen as unfair. In future work, it would be interesting to explore how DP impacts applicants and is perceived when combined with a different type of approach to credit policies that prevents increase in inequality.

8.4.6 Themes that cut across different studies

As this work took an exploratory and inductive approach the research team was open to findings that might not be related to the specific studies' aims. There were a few themes identified across the different research activities which were not related to the research questions defined: lender regulator balance, transparency and agency. While these topics are not related to DP they reflect the state of the industry in today's world and potential for its change.

The Wonga case (see Chapter 2.1.) [74, 51] was often mentioned by both industry and regulators as well as by applicants/users, and relates to the balance between lenders and regulators. Industry and Regulator used it as an example of the regulatory framework working well, as well as a "bad apple" example justifying the need for regulation (Chapter 5.3.). From the perspective of consumers (Chapter 7) it was discussed related to unfair credit policies and their impact on the consumers. It is further used as an example to show agreement with regulation and cap to interest rates.

From the part of the consumers, it seems like there is still some unresolved tension regarding the industry balance, i.e. agreement with application processes and regulatory workings but some level of disagreement with credit policies, specifically risk-based pricing. This lack of balance identified might be especially felt by consumers due to the current cost of living [86], leading to a consumer reflection on how the industry's current workings impacts people.

Also from the perspective of the consumer, the transparency and agency topics arose in Chapter 4 when discussing participants' knowledge and confidence of the application process. Most had a general idea however, they were not very confident about this knowledge. Furthermore, participants expressed a desire to have more information on how each data source is used in the decision process, generally more detailed information on the process as well as explanations on outcomes, especially if someone had been rejected. Having more detailed information, especially the explanation of outcomes would allow consumers the choice and opportunity to improve their financial behaviour and make changes in the future. However, from the perspective of the industry, outcome explanations is seen as enabling the consumers to "game" the system (which transpires that the lender can view consumers as being ill-intended, over having a more empathetic per-

spective). In the eyes of the industry, based on the findings of Chapter 5, an explanation of outcomes would be breaching institutions' intellectual property and risk losing their competitive advantage especially when addressing credit policies (where lenders tend to be a bit more open regarding the technologies used).

This leads to the question of what would increase transparency do to the industry?

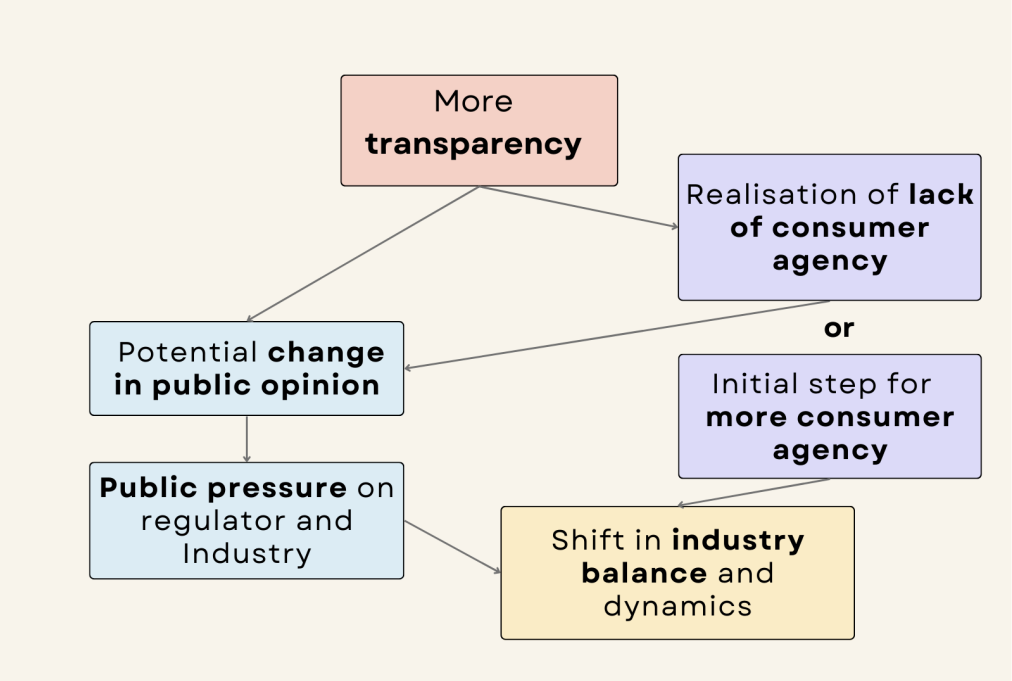


Figure 8.1: Transparency, agency and balance.

Figure 8.1 briefly summarises what could happen if there was an increase in transparency over the application process. If consumers did not agree with the workings of the process this would lead to:

1. a change in public opinion then putting pressure on the regulator and industry to address this

2. or if there was more transparency and the public perception didn't change it would further engrain the current processes

DP, and all other tech can also shift the relationship between transparency, industry balance and agency. Specifically, due to DP's post-processing property, it would be possible to have private outcome explanations, which could help consumers' financial lives. Currently, the industry does not seem to have the drive for such thing. It is understandable lenders reluctance to share some information, due to their competitive nature. However, the same cannot be said for the regulator, which I argue should be fighting for more transparency and agency for the consumer as it is part of consumer impact, the regulator's main aim. The regulator might not have motivation or an incentive to explore these alternatives as there is no public pressure , it would involve enagage with stakeholders specialised with DP without an immediate and guaranteed return, or they might not be aware of this potential implication of DP.

8.5 Potential Repercussions of DP Implementation in Risk Assessment Models of Loan Applications

This section is the culmination of the work of the thesis. It creates a more complete picture of the possible impact of implementing DP in Risk Assessment Models of loan applications in the consumer credit industry by triangulating the findings of the four research activities.

RQ1 and RQ2 served to gather the initial information on loan application process- from the perspectives of the consumer and the industry.

RQ4 was based on the findings from RQ2 (the choice of decision tree algorithms due to its commonplace implementation in the industry). RQ4

focuses on understanding the behaviour of the technology with financial daatsets.

RQ3 and RQ5 explore the impact of DP within the loan process. By discussing hypothetical scenarios with the Industry (to answer RQ3) based on different privacy-accuracy trade-off behaviours of DP and combining those findings with findings from RQ2 it was possible to start understanding how DP could affect the loan application process. This information was then combined with the findings based on RQ1 and was used to summarise and create a series of scenarios via a game board to explore consumers' attitudes towards DP within the loan application process (to answer RQ5).

In this section we will combine the findings from RQ 3,4,5 to understand the potential implications and impact of implementing DP in the loan application process.

8.5.1 What might happen if DP was implemented?

The impact of DP and attitudes towards its implementation would change considerably depending on the specific privacy-accuracy trade-off behaviour present.

The industry would be focused on loss of model's performance and profit. Consumers and the regulator would be focused on the impact to access to credit.

The different cases presented in this section describe what would happen if DP was implemented based on the findings from RQ 3,4 and 5. Each case study differs as it has a different privacy-accuracy trade-off behaviour.

Case 1: negligible accuracy drop

Based on the findings from Chapter 6, if the DP model implemented behaved similarly to DPGBDT 1x100, the risk score model would have a negligible accuracy drop and no DAL or ODAL (Chapter 6). Therefore, applicant's risk scores would be very similar to their non-private risk scores, meaning that they would probably have access to similar credit products. This could change for applicants in boundary values of getting credit or not, similar to situations in DP loan game of Chapter 7 if the credit policy remained the same.

As stated in Chapter 5, the risk and credit policy components are built separately, but the credit policy accounts for external risk factors such as macroeconomics. Lenders could adjust to the DP implementation by increasing the exogenous risk (external factors that get taken into consideration) in the NPV models which lead to the credit policies (see Chapter 5). However, credit policy tends to be built separately from risk assessment models.

Due to the principle-based approach to regulation, if the combination of accuracy changes and credit policy (if changed or not) is shown not to affect applicants access to affordable credit the regulators would not oppose it, as they are mainly concerned with consumer impact over model implementations, based on findings from Chapter 5.

If the accuracy drop was minimal to the point that its impact on consumers was negligible the lender would not oppose its implementation. If there was no profit loss lenders would still not likely implement it as the benefits to them would be minimal (minimizing reputational risk, better data protection) compared to the cost (training personnel, creation of a different risk

model), unless there was strong encouragement by the regulator.

If the impact on consumers' access to credit was very small and they were made aware of it, consumers might more easily accept the implementation of DP. However, due to the findings from this research, it is likely that consumers will likely still see the small addition of randomness as unfair.

Considering subgroup accuracy drop behaviour The amplitude of the accuracy drop across different subgroups would be the one of the determining factors in its impact. If there was some DAL, by a small degree, it might still not affect the subgroup's access to credit, depending on how close to the cut-off from not having access to any products they are. The impact of ODAL in terms of loan allocation is not as impactful unless the changes to the credit policy are significant.

Based on the findings from Chapter 7, we also know that consumers would find it more acceptable if those most affected by a subgroup disparate accuracy drop would be those with already better access to credit and in a good financial situation (as the "actual" impact on those consumers lives would be smaller compared to other groups). Furthermore, in practice if the technology was implemented consumers would probably not be given detailed information on how it would impact different subgroups (Chapter 5).

Case 2: significant accuracy drop

On the other hand, if the accuracy drop would be significant neither the lender (due to reduction of profits) nor the regulator (due to consumer impact) would want it to be implemented.

If the accuracy drop due to DP was more significant it could potentially

lead to big changes in credit policy (more conservative) hence decreasing access to credit. This would go against regulatory aims as well as going against the expansive character of the credit industry, described in Chapter 5. Alternatively, if the credit policy was not adjusted then more applicants might be allocated loans they could not afford or be denied access to loans they could afford. This would decrease lender's profit and negatively impacting consumers.

Considering subgroup accuracy drop behaviour If there are big discrepancies amongst subgroups, especially if they differ from non-private models, regulators might oppose the implementation. However, currently, lenders can target different subgroups of the population, hence having discriminatory credit policies. In the stakeholder interview study (Chapter 5) the topics of algorithmic fairness and accuracy differences for subgroups was discussed with a couple of participants. They stated that accuracy equality was not something that was pursued by the industry. As such they were not certain how regulator might react but one could predict the consumer to perceive this in a negative light depending on who is impacted the most.

8.5.2 How feasible is DP to be implemented in the risk assessment model of the UK consumer credit industry?

From the data collected in the research activities presented in the thesis, it does not seem very feasible for DP to be implemented.

From the lender's perspective, the loss of competitiveness, if not all lenders implement the technology, deters its implementation. Even if they were

interested in implementing DP, there would still exist the first mover problem, so the regulator would need to show some interest to motivate this.

From the regulatory perspective there will not be a push due to the potential negative impact to the consumer as even with a small accuracy drop there are not enough positives to outweigh this. However, if there is a change in public attitude, especially if due to a scandal or data leak that could have been prevented with DP, regulators might be motivated to act.

8.6 Main Contributions

Figure 8.2 summarises the main contributions of this thesis. The figure further highlights which stakeholders these contributions are relevant to and where in the thesis the findings that support it are.

Thesis Contributions (non-extensive)	
1. New ways to communicate DP	
Summary: Design of the glasses analogy and interactive game to communicate DP **Who is it useful to:** • people implementing DP (range of industries and government departments) • DP academics (especially those working within Usable DP) • consumer - better understanding of what happens to their data with DP	**Findings in:** • Chapter 1.1 • Chapter 7.2
2. New data that shows consumer desire for more transparency in the credit industry	
Summary: Participants express wanting more information regarding how their data is used in the application process, more details regarding the process and explanations of outcomes. **Who is it useful to:** • Credit industry - could lead to improvement of consumer relationships • Regulator (FCA)	**Findings in:** • Chapter 4.3 • Chapter 7.3
3. New knowledge on the privacy-accuracy trade-off of DP Decision Tree models	
Summary: DP- Gradient Boosting Decision Tree and Smooth Random Forest models have comparable performances to non-private models and no significant disparate accuracy loss for data subgroups. **Who is it useful to:** • people implementing DP (range of industries and government departments) • DP academics	**Findings in:** • Chapter 6.3
4. New knowledge on consumer attitudes towards DP (and its varying privacy accuracy trade-off behaviours) in the consumer credit industry	
Summary: Varied views on prefered DP behaviour and random element associated with DP is seen as unfair in the loan application setting. **Who is it useful to:** • Credit industry - could lead to improvement of consumer relationships • Regulator (FCA) • DP academics (especially those working within Usable DP)	**Findings in:** • Chapter 7.3
5. New knowledge on the potential impact of DP implementation on the credit industry	
Summary: DP would probably not negatively impact applicants disparately, however access to credit might become more restricted due to changes to credit policy **Who is it useful to:** • Credit Industry • Regulator • Consumers • DP academics	**Findings in:** • Chapter 5.3 • Chapter 6.3 • Chapter 7.3

Figure 8.2: Major Contributions Summary Table

The order in which the contributions are shown in Figure 8.2 corresponds to how immediate they are to real world applications.

- **New ways to communicate DP**

As a result of the design of the Industry stakeholder and the focus group studies I generated different ways to communicate DP.

The glasses analogy showcases how DP is able to gather aggregated information while maintaining individual privacy by the addition of noise. It also represents the privacy parameters/noise and its relation to accuracy via the amount of fog present, fulfilling the call for more accurate explanations in Smart [150].

The interactive game board was able to transmit knowledge about the technology in an accessible and inclusive manner via the game play. The different dice elements and their role in the game play was used to simulate and explain the way randomness works within DP, which is at times challenging to comprehend [167] and the comparison between different types of models helps contextualize the impact of them in relation to each other, an approach based on the findings of Nanayakkare et al [117].

These two methods of communicating DP can be either altered and/or immidiately employed by the variety of institutions already implementing the technology. As a result, these methods are also of relevance to consumers- they provide an attempt at an active way of learning (hands-on) and understanding about a technology that is often discussed in overly technical terms. By improving understanding of how the technology works, this can lead to consumers being more informed and understand the potential impact it can have in their data. In the case of using DP as a way to get consumers to feel more comfortable to share a wider range of personal data, as often tested in the Usable DP field, understanding accurately how the technology works is essential to informed decision making.

Finally, these two DP explanations are of use to Usable DP academics, who face the challenge of communicating DP in an accessible way. Use

of the analogy and game should be tested for efficacy of understanding of concepts and be compared to a variety of other methods already presented in the literature.

Of particular importance might be the game as a novel way of interactive explanations, hence it could be adapted to other implementation scenarios and tested. The game acts as an example of how FinTech could be explored and communicated to diverse consumers, similarly to the video games presented in [136, 139] which focus on young adults and social workers respectively. So this PhD has shown success in its application, other games could be developed to improve financial literacy, which will be discussed in more detail in the Future Work section of this chapter.

- **New data which shows consumer desire for more transparency in the credit industry**

Some of the findings of Chapter 4 highlighted a consumer desire to have access to more detailed information about the application process, its decision making and how consumer's personal data is used within that process.

This data showcases that the effort the credit industry has been making to improve consumer trust as a result of the 2008 financial crisis [4] has not been fully accomplished. As such, this data should be of interest to the credit industry. It provides requirements of the banking institutions to improve communications transparency regarding their decision processes. Furthermore, this data could potentially be of importance to the FCA (UK consumer credit regulator), as their main driving factor is consumer protection as per their mission statement [11]. The data could serve as evidence for the need of communications targeted at the consumer to fill this knowledge gap. The regulator could also encourage the industry to

improve communication and reduce opacity, for the benefit of the industry as it might lead to an increase in consumer trust and to address the requirements expressed by participants in this research.

- **New knowledge on the privacy-accuracy trade-off of DP Decision Tree models**

The implementation of a range of different decision tree based models and its analysis of the differing overall performance and privacy-accuracy trade-off for different dataset subgroups provides important information for those looking to implement such family of models, such as data scientists in both public and private organisations.

It gives insights on how the models react to different types of datasets and through the privacy- accuracy graphs it indicates how to start choosing a privacy parameter for each of the models.

This knowledge expands on the work of Fletcher et al [68] by analysing the impact of the accuracy drop (at a variety of different privacy parameters) for different subgroups within the training datasets. As such this knowledge is also of use to the DP academic community, who could analyse the implemented algorithms from a theorical perspective.

- **New knowledge on consumer attitudes towards DP**

As a result of the focus group research findings discussed in Chapter 7, new knowledge on consumers' attitudes towards DP was created.

This knowledge accounts for different privacy-accuracy trade-off behaviours, with different models having either an equal impact in different subgroups or a differing one. This is useful to Usable DP experts in academia, as it is

an extension and more in depth analysis of factors which affect attitudes towards differing DP models. The work is also relevant to industry experts which are implementing DP or considering it in the future.

As the work is also set within the specific setting of a loan application process the knowledge created is also of use to both the regulator and the consumer credit industry as it can give an indication of consumers' attitudes previous to implementation of DP (if such would be considered).

As stated earlier, there is a growing importance on consumers' usage and acceptance of technology [101]. Against this backdrop, the findings provide new knowledge to the regulators to support this goal.

- **New knowledge on the potential impact of DP implementation in the credit industry**

Resulting from the triangulation of findings previously reported in the chapter this PhD has generated new knowledge on the potential impact of DP implementation. At the moment this new knowledge is of most use to industry and regulators, helping shape the decision making regarding the implementation of this technology as explained in section 8.3.3. In the case of inplementation it would be of further usel to consumers as it relates to the processing of their personal data and has an impact on their financial lives.

This work is also relevant to academics working with DP as it is the first study of the technology accounting for its deployment setting, in the thesis case the risk assessment model in uk consumer credit loan applications, and investigating its possible impacts to both consumers and the industry. This work sets an approach to the sociotechnical study of the technology which should be built upon in future work.

- **Other contributions**

The game board designed for the Differentially Private Consumer Credit Imaginaries study (Chapter 7) besides serving as an interactive explanation of DP it also serves as a high-level explanation of the workings of the loan application process.

In such a complex landscape in which it is well established that financial literacy of the general public is poor [35, 76] and industry processes and associated technology are ever more complex, the game board serves as a tool to engage and expose the general public in an accessible and interactive

manner. The game has the potential to be adapted for financial literacy education, which has used games for education in a variety of financial scenarios [20, 138].

8.7 Future Work

This research has delivered understanding in the Usable DP field, however, there is a significant opportunity for future research to further explore different applied scenarios within and outside of the credit industry and their associated sociotechnical systems.

Firstly, the research established that consumers have diverse needs and preferences towards the mechanisms and decisioning of the loan application process. Further work could explore how to design a variety of products and processes based on consumers' needs. This would be useful because of the vital and impactful role the credit industry can play in people's lives. Taking a bottoms-up approach by starting to gather consumers' needs and requirements regarding credit products and using this to design them, would be a novel approach within the industry, and could help address some of their needs discussed in Chapter 4. It would also be possible to understand how different products and policies could affect different sections of the population by altering and using a game board such as the one described in Chapter 7.

As focus group participants reacted well to the personas created, it might be benefitial to develop a personas deck, with more variety and detail. While other persona decks focused on the financial industry might exist to the best of my knowldege they are not open access. The personas deck could help researcher investigating consumers' attitudes towards the financial industry

and its technology.

We now understand more about the accuracy loss behaviour of differentially private decision-based tree models. However, there is still limited knowledge of this behaviour on non-financial datasets. As there seems to be a correlation between the type of covariates and model performance, in different contexts with other datasets the models tested might behave very differently, hence further work should be pursued.

Novel technology in this space moves quickly. It would be valuable to develop novel research methods and tools to be able to effectively investigate technologies' impact on sociotechnical systems. This development of the game in the interactive focus group study provides a new way of examining user attitudes towards DP in a loan application context. This needs further testing and iterative design to improve DP communication and knowledge acquisition. Only once user mental models of DP are in line with the mechanisms and privacy implications of the technology they can make informed decisions about under which circumstances they would like DP to be deployed.

In addition, more work is needed to design and test methods that make this complex system/ service accessible and understandable to all. Developing methods that are accessible and thereby improving understanding of consumer needs also feeds into the Woolard review's goals [172] of developing financial technologies that are fair and trustworthy. In order to achieve this goal, it is essential to involve diverse stakeholders.

8.8 Conclusion

Differential Privacy is currently seen as the state-of-the-art PET [126], having been employed in several industrial (e.g. Google, facebook, LinkedIn) and governmental contexts (US Census) [175]. This technology has an associated accuracy-privacy trade-off. DP is a very technical definition of privacy hence complicated to communicate with the general public. Furthermore, some of its models have been shown to have disparate effects on subgroups of the population. This PhD project was designed to understand the potential repercussions of the implementation of DP within the risk assessment model for consumer credit focusing on a consumer perspective.

In order to achieve this goal a variety of research methods were employed, with different research studies focused on the different stakeholders: consumer exploratory interviews on the application process, industry stakeholder consultation, implementation and comparison of different differentially private decision tree-based algorithms and a group game-based interactive study to gather consumers' attitudes towards the implementation of the technology.

From the technical study, it was found that certain algorithms had a neglectable accuracy drop for privacy budgets bigger than 4 when compared to a non-private algorithm and rare occasions of disparate accuracy loss. Combining this information with the findings from the Industry Consultation (Chapter 4) tells us that if DP was implemented the majority of consumers would not be significantly affected (mainly the ones near the cut-off point regarding having access to any type of credit or not) and businesses would not be as impacted. To compensate for the implementation, lenders could change their credit policy to account for the small increase in uncertainty in the risk scores making credit less accessible, which goes against

regulatory aims. Based on findings from Chapter 7, consumers have very mixed views regarding the implementation of DP, as they would rather have better financial options than protect their personal data. Based on these combined findings DP is unlikely to the implemented as lenders would require some regulatory encouragement which seems unlikely unless there is a shift in public opinion.

This work contributes to the underrepresented area of usable DP and the sociotechnical approach to DP. Furthermore, this work sheds light on the opaque credit industry and the consumer desire for more transparency of processes and outcomes.

Bibliography

[1] callforparticipants.com. https://www.callforparticipants.com/. [Accessed 21-08-2023].

[2] Credit card advert, 1972 — NatWest Group Heritage Hub — natwestgroup.com. https://www.natwestgroup.com/heritage/history-100/objects-by-theme/serving-our-customers/credit-card-advert-1972.html#:~:text=It%20was%20not%20until%201966,venture%2C%20and%20called%20it%20Access. [Accessed 20-Apr-2023].

[3] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[4] Shakeel Ahmed, Kenbata Bangassa, and Saeed Akbar. A study on trust restoration efforts in the uk retail banking industry. *The British Accounting Review*, 52(1):100871, 2020.

[5] Abdullah Alanazi. Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, page 100924, 2022.

[6] Saleema Amershi, James Fogarty, and Daniel Weld. Regroup: Interactive machine learning for on-demand group creation in social

networks. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 21–30, 2012.

[7] Mandy M Archibald, Rachel C Ambagtsheer, Mavourneen G Casey, and Michael Lawless. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. *International journal of qualitative methods*, 18:1609406919874596, 2019.

[8] Aycan Aslan, Tizian Matschak, Maike Greve, Simon Trang, and Lutz Kolbe. At what price? exploring the potential and challenges of differentially private machine learning for healthcare. 2023.

[9] Paul Atkinson. *Ethnography: Principles in practice*. Routledge, 2007.

[10] Financial Conduct Authority. Prin 2.1 the principles. https://www.handbook.fca.org.uk/handbook/PRIN/2/1.html. [Accessed 23-Apr-2023].

[11] Financial Conduct Authority. Our future mission. https://www.fca.org.uk/news/press-releases/fca-mission-consultation, 2016. [Accessed 23-Apr-2023].

[12] Financial Conduct Authority. Financial lives 2020 survey: the impact of coronavirus. https://www.fca.org.uk/publication/research/financial-lives-survey-2020.pdf, 2021. [Accessed 1-Jan-2025].

[13] Financial Conduct Authority et al. Machine learning in uk financial services. https://www.fca.org.uk/publication/research/research-note-on-machine-learning-in-uk-financial-services.pdf, 2019. [Accessed 24-Apr-2023].

[14] Eugene Bagdasaryan, Omid Poursaeed, and Vitaly Shmatikov. Dif-

ferential privacy has disparate impact on model accuracy. *Advances in Neural Information Processing Systems*, 32:15479–15488, 2019.

[15] Eugene Bagdasaryan and Vitaly Shmatikov. Differential Privacy Has Disparate Impact on Model Accuracy. *arXiv:1905.12101 [cs, stat]*, October 2019. arXiv: 1905.12101.

[16] Janine Baleis, Birte Keller, Christopher Starke, and Frank Marcinkowski. Cognitive and Emotional Response to Fairness in AI – A Systematic Review. page 42.

[17] Janine Baleis, Birte Keller, Christopher Starke, and Frank Marcinkowski. Cognitive and emotional response to fairness in ai–a systematic review. 2019.

[18] Borja Balle and Yu-Xiang Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 394–403. PMLR, 10–15 Jul 2018.

[19] Prudential Regulation Authority Bank of England, Financial Conduct Authority. DP5/22 - Artificial Intelligence and Machine Learning — bankofengland.co.uk. https://www.bankofengland.co.uk/prudential-regulation/publication/2022/october/artificial-intelligence. [Accessed 24-Apr-2023].

[20] Julia Bayuk and Suzanne Aurora Altobello. Can gamification improve financial behavior? the moderating role of app expertise. *International Journal of Bank Marketing*, 37(4):951–975, 2019.

[21] Allison Benedetti, John Jackson, and Lili Luo. Vignettes: Implications for lis research. *College & Research Libraries*, 79(2):222, 2018.

[22] Wicher Bergsma. A bias-correction for cramér's v and tschuprow's t. *Journal of the Korean Statistical Society*, 42(3):323–328, 2013.

[23] David Bholat, Mohammed Gharbawi, and Oliver Thew. The impact of covid on machine learning and data science in uk banking. *Bank of England Quarterly Bulletin*, page Q4, 2020.

[24] Yochanan E Bigman and Kurt Gray. People are averse to machines making moral decisions. *Cognition*, 181:21–34, 2018.

[25] Ann Blandford, Dominic Furniss, and Stephann Makri. *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers, 2016.

[26] Amanda Bolderston. Conducting a research interview. *Journal of medical imaging and radiation sciences*, 43(1):66–76, 2012.

[27] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[28] Virginia Braun and Victoria Clarke. Toward good practice in thematic analysis: Avoiding common problems and be (com) ing a knowing researcher. *International Journal of Transgender Health*, 24(1):1–6, 2023.

[29] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[30] Alan Bryman. *Social research methods*. Oxford university press, 2016.

[31] Alan Bryman, Saul Becker, and Joe Sempik. Quality criteria for quantitative, qualitative and mixed methods research: A view from

social policy. *International journal of social research methodology*, 11(4):261–276, 2008.

[32] Bonnie G Buchanan and Danika Wright. The impact of machine learning on uk financial services. *Oxford Review of Economic Policy*, 37(3):537–563, 2021.

[33] Brooke Bullek, Stephanie Garboski, Darakhshan J Mir, and Evan M Peck. Towards understanding differential privacy: When do people trust randomized response technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3833–3837, 2017.

[34] Hanna Carlsson, Stefan Larsson, Lupita Svensson, and Fredrik Åström. Consumer credit behavior in the digital context: A bibliometric analysis and literature review. *Journal of financial counseling and planning*, 28(1):76–94, 2017.

[35] Hanna Carlsson, Stefan Larsson, Lupita Svensson, and Fredrik Åström. Consumer Credit Behavior in the Digital Context: A Bibliometric Analysis and Literature Review. *Journal of Financial Counseling and Planning*, 28(1):76–94, 2017.

[36] Peter Checkland and Alejandro Casar. Vickers' concept of an appreciative system: a systemic account. *Journal of Applied Systems Analysis*, 13(3):3–17, 1986.

[37] Michael Chui, Bryce Hall, Hellen Mayhew, Alex Singla, Alex Sukharevsky, and AI by McKinsey. The state of ai in 2022-and a half decade in review. *https://www.mckinsey. com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/*, 2022.

[38] Alexander M Clark. The qualitative-quantitative debate: moving from positivism and confrontation to post-positivism and reconciliation. *Journal of advanced nursing*, 27(6):1242–1249, 1998.

[39] Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. *Qualitative psychology: A practical guide to research methods*, 222(2015):248, 2015.

[40] John C Coffee Jr. What went wrong? an initial inquiry into the causes of the 2008 financial crisis. *Journal of Corporate Law Studies*, 9(1):1–22, 2009.

[41] Vinicius G Costa and Carlos E Pedreira. Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5):4765–4800, 2023.

[42] Kelley Cotter. Practical knowledge of algorithms: The case of breadtube. *new media & society*, page 14614448221081802, 2022.

[43] Kathrin M Cresswell and Aziz Sheikh. Undertaking sociotechnical evaluations of health information technologies. *Journal of Innovation in Health Informatics*, 21(2):78–83, 2014.

[44] John W Creswell and Cheryl N Poth. *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, 2016.

[45] J.W. Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 2014.

[46] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. " i need a better description": An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 3037–3052, 2021.

[47] Gerald Cupchik et al. Constructivist realism: An ontology that encompasses positivist and constructivist approaches to the social sciences. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, volume 2, 2001.

[48] Xavier De Scheemaekere. The epistemology of modern finance. *Journal of Philosophical Economics*, 2(Articles), 2009.

[49] Laura Dempsey, Maura Dowling, Philip Larkin, and Kathy Murphy. Sensitive interviewing in qualitative research. *Research in nursing & health*, 39(6):480–490, 2016.

[50] Elizabeth Depoy and Laura Gitlin. *Collecting Data Through Measurement in Experimental-Type Research*, pages 227–247. 12 2016.

[51] Joe Deville. Digital subprime: Tracking the credit trackers. In *The sociology of debt*, pages 145–174. Policy Press, 2019.

[52] Joe Deville. Futures of credit risk assessment in the uk. 2020.

[53] Owen Doody and Maria Noonan. Preparing and conducting interviews to collect data. *Nurse researcher*, 20(5), 2013.

[54] Maureen Duffy. Sensemaking: A collaborative inquiry approach to" doing" learning. *The Qualitative Report*, 2(2):1–6, 1995.

[55] John Durham Peters. You mean my whole fallacy is wrong: On technological determinism. *You Mean My Whole Fallacy is Wrong: On Technological Determinism*, pages 26–34, 2019.

[56] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality*, 9(2), October 2019.

[57] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[58] Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.

[59] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[60] Laura Cox Dzurec. The necessity for and evolution of multiple paradigms for nursing research: A poststructuralist perspective. *Advances in Nursing Science*, 11(4):69–77, 1989.

[61] Peter Edwards, Lisa Sharma-Wallace, Anita Wreford, Lania Holt, Nicholas A Cradock-Henry, Stephen Flood, and Sandra J Velarde. Tools for adaptive governance for complex social-ecological systems: a review of role-playing-games as serious games at the community-policy interface. *Environmental Research Letters*, 14(11):113002, 2019.

[62] Liyue Fan. A survey of differentially private generative adversarial networks. In *The AAAI Workshop on Privacy-Preserving Artificial Intelligence*, 2020.

[63] Wei Fan, Haixun Wang, Philip S Yu, and Sheng Ma. Is random model better? on its accuracy and efficiency. In *Third IEEE International Conference on Data Mining*, pages 51–58. IEEE, 2003.

[64] Tom Farrand, Fatemehsadat Mireshghallah, Sahib Singh, and Andrew Trask. Neither Private Nor Fair: Impact of Data Imbalance

on Utility and Fairness in Differential Privacy. In *Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice*, PPMLP'20, pages 15–19, New York, NY, USA, November 2020. Association for Computing Machinery.

[65] Todd Fernandez, Allison Godwin, Jacqueline Doyle, Dina Verdin, Hank Boone, Adam Kirn, Lisa Benson, and Geoff Potvin. More comprehensive and inclusive approaches to demographic data collection. 2016.

[66] Sam Fletcher and Md Zahidul Islam. A differentially private random decision forest using reliable signal-to-noise ratios. In *Australasian joint conference on artificial intelligence*, pages 192–203. Springer, 2015.

[67] Sam Fletcher and Md Zahidul Islam. Differentially private random decision forests using smooth sensitivity. *Expert systems with applications*, 78:16–31, 2017.

[68] Sam Fletcher and Md Zahidul Islam. Decision tree classification with differential privacy: A survey. *ACM Computing Surveys (CSUR)*, 52(4):1–33, 2019.

[69] Marc Fleurbaey. Equal opportunity or equal social outcome? *Economics & Philosophy*, 11(1):25–55, 1995.

[70] Brigid Francis-Devine. Which ethnic groups are most affected by income inequality. *Commons Library*, 10(August), 2020.

[71] Arik Friedman and Assaf Schuster. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–502, 2010.

[72] Frederik Funke and Ulf-Dietrich Reips. Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field methods*, 24(3):310–327, 2012.

[73] Nicola K Gale, Gemma Heath, Elaine Cameron, Sabina Rashid, and Sabi Redwood. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC medical research methodology*, 13(1):1–8, 2013.

[74] Nicholas Gane. Debt, usury and the ongoing crises of capitalism. In *The sociology of debt*, pages 175–194. Policy Press, 2019.

[75] Georgi Ganev, Bristena Oprisanu, and Emiliano De Cristofaro. Robin hood and matthew effects: Differential privacy has disparate impact on synthetic data. In *International Conference on Machine Learning*, pages 6944–6959. PMLR, 2022.

[76] John Gathergood and Richard F Disney. Financial literacy and indebtedness: new evidence for uk consumers. *Available at SSRN 1851343*, 2011.

[77] Damon Gibbons. *Britain's Personal Debt Crisis: How we got here and what to do about it.* Searching Finance, 2014.

[78] John W Goodell, Satish Kumar, Weng Marc Lim, and Debidutta Pattnaik. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, 32:100577, 2021.

[79] GOV.UK. Income tax liabilities statistics: Tax year 2017 to 2018, to tax year 2020 to 2021. www.gov.uk/government/statistics. [Accessed 16-Nov-2022].

[80] GOV.UK. Population of england and wales. www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest. [Accessed 24-Apr-2023].

[81] Judith A Greene. Science, nursing and nursing science: A conceptual analysis. *Advances in Nursing Science*, 2(1):57–64, 1979.

[82] Ziwei Gu, Jing Nathan Yan, and Jeffrey M Rzeszotarski. Understanding user sensemaking in machine learning fairness assessment systems. In *Proceedings of the Web Conference 2021*, pages 658–668, 2021.

[83] Egon G Guba. The alternative paradigm dialog. in. eg guba.(ed). the paradigm dialog, 1990.

[84] Madeline J Hallewell, Nancy Hughes, David R Large, Catherine Harvey, James Springthorpe, and Gary Burnett. Deriving personas to inform hmi design for future autonomous taxis: A case study on user requirement elicitation. *Journal of Usability Studies*, 17(2), 2022.

[85] Karin Hammarberg, Maggie Kirkman, and Sheryl de Lacey. Qualitative research methods: when to use them and how to judge them. *Human reproduction*, 31(3):498–501, 2016.

[86] Daniel Harari, Brigid Francis-Devine, Paul Bolton, and Matthew Keep. Rising cost of living in the uk. *London: House of Commons Library https://commonslibrary. parliament. uk/research-briefings/cbp-9428*, 2022.

[87] Lois R Harris and Gavin TL Brown. Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment, Research, and Evaluation*, 15(1):1, 2019.

[88] Mohammad Hossin and Md Nasir Sulaiman. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1, 2015.

[89] Xixi Huang, Ye Ding, Zoe L Jiang, Shuhan Qi, Xuan Wang, and Qing Liao. Dp-fl: a novel differentially private federated learning framework for the unbalanced data. *World Wide Web*, 23(4):2529–2545, 2020.

[90] Rhidian Hughes and Meg Huby. The application of vignettes in social and nursing research. *Journal of advanced nursing*, 37(4):382–386, 2002.

[91] Mikella Hurley and Julius Adebayo. Credit scoring in the era of big data. *Yale JL & Tech.*, 18:148, 2016.

[92] Phillip Inman. Consumer credit races ahead as UK households struggle to cope. https://www.theguardian.com/business/2022/jul/29/consumer-credit-races-ahead-as-uk-households-struggle-to-cope, 2022. [Accessed 24-Apr-2023].

[93] Christine Ironfield-Smith, Kevin Keasey, Barbara Summers, Darren Duxbury, and Robert Hudson. Consumer debt in the uk: Attitudes and implications. *Journal of Financial Regulation and Compliance*, 2005.

[94] Geetha Jagannathan, Krishnan Pillaipakkamnatt, and Rebecca N Wright. A practical differentially private random decision tree classi-

fier. In *2009 IEEE International Conference on Data Mining Workshops*, pages 114–121. IEEE, 2009.

[95] Mimansa Jaiswal and Emily Mower Provost. Privacy enhanced multimodal neural representations for emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7985–7993, 2020.

[96] Justin Joque. *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*. Verso Books, 2022.

[97] Natascha Just and Michael Latzer. Governance by algorithms: reality construction by algorithmic selection on the internet. *Media, culture & society*, 39(2):238–258, 2017.

[98] Farzaneh Karegar and Simone Fischer-Hübner. Vision: A noisy picture or a picker wheel to spin? exploring suitable metaphors for differentially private data analyses. In *Proceedings of the 2021 European Symposium on Usable Security*, pages 29–35, 2021.

[99] Satveer Kaur-Gill and Mohan J Dutta. Digital ethnography. *The international encyclopedia of communication research methods*, 10(1), 2017.

[100] Sam Keen, Martha Lomeli-Rodriguez, and Helene Joffe. From challenge to opportunity: virtual qualitative research during covid-19 and beyond. *International Journal of Qualitative Methods*, 21:16094069221105075, 2022.

[101] Mirella Kleijnen, Martin Wetzels, and Ko De Ruyter. Consumer acceptance of wireless finance. *Journal of financial services marketing*, 8:206–217, 2004.

[102] Gunther Kress. Critical discourse analysis. *Annual review of applied linguistics*, 11:84–99, 1990.

[103] Patrick Kühtreiber, Viktoriya Pak, and Delphine Reinhardt. Replication: The effect of differential privacy communication on german users' comprehension and data sharing attitudes. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 117–134, 2022.

[104] Fedwa Laamarti, Mohamad Eid, and Abdulmotaleb El Saddik. An overview of serious games. *International Journal of Computer Games Technology*, 2014:11–11, 2014.

[105] Tony Lawson. Reorienting economics: on heterodox economics, themata and the use of mathematics in economics. *Journal of economic methodology*, 11(3):329–340, 2004.

[106] Min Kyung Lee. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1):2053951718756684, 2018.

[107] Qinbin Li, Zhaomin Wu, Zeyi Wen, and Bingsheng He. Privacy-preserving gradient boosting decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 784–791, 2020.

[108] Yvonna S Lincoln, Susan A Lynham, Egon G Guba, et al. Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research*, 4(2):97–128, 2011.

[109] Xiaoqian Liu, Qianmu Li, Tao Li, and Dong Chen. Differentially private classification with decision tree ensemble. *Applied Soft Computing*, 62:807–816, 2018.

[110] Bojana Lobe, David Morgan, and Kim A Hoffman. Qualitative data collection in an era of social distancing. *International journal of qualitative methods*, 19:1609406920937875, 2020.

[111] Sigrun Lurås. Systems intertwined: a systemic view on the design situation. *Design issues*, 32(3):30–41, 2016.

[112] Eva Magnusson and Jeanne Marecek. *Doing the interview*, page 58–72. Cambridge University Press, 2015.

[113] Philipp Mayring. Qualitative content analysis, forum. *Qualitative social research*, 1(2):1–10, 2000.

[114] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[115] Noman Mohammed, Rui Chen, Benjamin CM Fung, and Philip S Yu. Differentially private data release for data mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 493–501, 2011.

[116] Shane T Mueller and Yin-Yin Sarah Tan. Cognitive perspectives on opinion dynamics: The role of knowledge in consensus formation, opinion divergence, and group polarization. *Journal of Computational Social Science*, 1:15–48, 2018.

[117] Priyanka Nanayakkara, Mary Anne Smart, Rachel Cummings, Gabriel Kaptchuk, and Elissa Redmiles. What are the chances? explaining the epsilon parameter in differential privacy. *arXiv preprint arXiv:2303.00738*, 2023.

[118] Alexey Natekin and Alois Knoll. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7:21, 2013.

[119] Lene Nielsen. A model for personas and scenarios creation. *COGNITIVE SCIENCE RESEARCH PAPER-UNIVERSITY OF SUSSEX CSRP*, pages 38–40, 2003.

[120] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[121] Donald A Norman. Some observations on mental models. In *Mental models*, pages 15–22. Psychology Press, 2014.

[122] Office of Fair Traiding. Irresponsible lending - OTF guidance for creditors. https://webarchive.nationalarchives.gov.uk/ukgwa/20140402161821mp_/http://oft.gov.uk/shared_oft/business_leaflets/general/oft1107.pdf. [Accessed 24-Apr-2023].

[123] Rina Okada, Kazuto Fukuchi, Kazuya Kakizaki, and Jun Sakuma. Differentially Private Analysis of Outliers. *arXiv:1507.06763 [cs, stat]*, July 2015. arXiv: 1507.06763.

[124] Karen O'Quin and Susan P Besemer. The development, reliability, and validity of the revised creative product semantic scale. *Creativity Research Journal*, 2(4):267–278, 1989.

[125] H Page, C Cabot, and K Nissim. Differential privacy an introduction for statistical agencies. *NSQR. Government Statistical Service*, 2018.

[126] Abhijit Patil and Sanjay Singh. Differential private random forest. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2623–2630. IEEE, 2014.

[127] Michael Quinn Patton. Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(5 Pt 2):1189, 1999.

[128] Balázs Pejó and Damien Desfontaines. Sok: Differential privacies. 2020.

[129] Trevor J Pinch and Wiebe E Bijker. The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social studies of science*, 14(3):399–441, 1984.

[130] Martha Poon. From new deal institutions to capital markets: Commercial consumer risk scores and the making of subprime mortgage finance. *Accounting, Organizations and Society*, 34(5):654–674, 2009.

[131] Roshani K Prematunga. Correlational analysis. *Australian Critical Care*, 25(3):195–199, 2012.

[132] J. Ross Quinlan. Learning decision tree classifiers. *ACM Computing Surveys (CSUR)*, 28(1):71–72, 1996.

[133] Syahirah Abdul Rahman, Lauren Tuckerman, Tim Vorley, and Cristian Gherhes. Resilient research in the field: Insights and lessons from adapting qualitative research projects during the covid-19 pandemic. *International journal of qualitative methods*, 20:16094069211016106, 2021.

[134] Oona Rainio, Jarmo Teuho, and Riku Klén. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1):6086, 2024.

[135] Chowdhury Raqib. Embarking on research in the social sciences: Understanding the foundational concepts. *VNU Journal of Foreign Studies*, 35(1), 2019.

[136] Aldrich Rasco, Johnny Chan, Gabrielle Peko, and David Sundaram. Fincraft: Immersive personalised persuasive serious games for finan-

cial literacy among young decision-makers. In *Proceedings of the 53rd Hawaii international conference on system sciences*, pages 32–41, 2020.

[137] Cheryl Jeanne Reifer. *Using focus group methodology to develop diabetes screening, education, and prevention programs for African American women*. Texas Woman's University, 2001.

[138] Lyudmyla Remnova and Khrystyna Shtyrkhun. Creative learning of finance and economics through gamification. *Teaching Methods for Economics and Business Sciences*, 2020.

[139] Kristin Richards, Jaclyn M Williams, Thomas E Smith, and Bruce A Thyer. Financial video games: A financial literacy tool for social workers. *International Journal of Social Work*, pages 22–35, 2015.

[140] Danya Rumore, Todd Schenk, and Lawrence Susskind. Role-play simulations for climate change adaptation education and engagement. *Nature Climate Change*, 6(8):745–750, 2016.

[141] Steven L Salzberg. C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993, 1994.

[142] Helen Sampson and Idar Alfred Johannessen. Turning on the tap: the benefits of using 'real-life'vignettes in qualitative research interviews. *Qualitative Research*, 20(1):56–72, 2020.

[143] Peter J Sandiford and John Ap. Important or not? a critical discussion of likert scales and likert-type scales as used in customer research. In *12th Annual CHME Hospitality Research Conference: Trend and developments in hospitality research, Sheffield Hallam University, Sheffield, South Yorkshire, England*, 2003.

[144] Jayshree Sarathy. From algorithmic to institutional logics: the politics of differential privacy. *Available at SSRN*, 2022.

[145] Steve Sawyer and Mohammad Hossein Jarrahi. Sociotechnical approaches to the study of information systems. In *Computing handbook, third edition: Information systems and information technology*, pages 5–1. CRC Press, 2014.

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

[147] Jakob Schoeffer, Yvette Machowski, and Niklas Kuehl. A study on fairness and trust perceptions in automated decision making. *arXiv preprint arXiv:2103.04757*, 2021.

[148] Donghee Shin and Yong Jin Park. Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98:277–284, 2019.

[149] Sergio Sismondo. *An introduction to science and technology studies*, volume 1. Wiley-Blackwell Chichester, 2010.

[150] Mary Anne Smart, Dhruv Sood, and Kristen Vaccaro. Understanding risks of privacy theater with differential privacy. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–24, 2022.

[151] Steven E Stemler. Content analysis. *Emerging trends in the social and behavioral sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pages 1–14, 2015.

[152] David Stokes and Richard Bergin. Methodology or "methodolatry"? an evaluation of focus groups and depth interviews. *Qualitative market research: An international Journal*, 9(1):26–37, 2006.

[153] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002.

[154] Hamed Taherdoost. How to conduct an effective interview; a guide to interview design in research study. *International Journal of Academic Research in Management*, 11(1):39–51, 2022.

[155] Behavioural Insights Team. The perception of fairness of algorithms and proxy information: A report for the centre for data ethics and innovation from the behavioural insights team in financial services., 2019.

[156] Matti Tedre. Know your discipline: Teaching the philosophy of computer science. *Journal of Information Technology Education: Research*, 6(1):105–122, 2007.

[157] Daniel Tischer, Bill Maurer, and Adam Leaver. Finance as 'bizarre bazaar': Using documents as a source of ethnographic knowledge. *Organization*, 26(4):553–577, 2019.

[158] Jukka Törrönen. Using vignettes in qualitative interviews as clues, microcosms or provokers. *Qualitative Research Journal*, 18(3):276–286, 2018.

[159] HM Treasury. Reform of the consumer credit act: Consultation, Jul 2023.

[160] Data Source Triangulation. The use of triangulation in qualitative research. In *Oncol nurs forum*, volume 41, pages 545–7, 2014.

[161] Gov UK. UK commits to reform of the Consumer Credit Act. https://www.gov.uk/government/news/uk-commits-to-reform-of-the-consumer-credit-act. [Accessed 24-Apr-2023].

[162] Archit Uniyal, Rakshit Naidu, Sasikanth Kotti, Sahib Singh, Patrik Joslin Kenfack, Fatemehsadat Mireshghallah, and Andrew Trask. Dp-sgd vs pate: Which has less disparate impact on model accuracy? *arXiv preprint arXiv:2106.12576*, 2021.

[163] Iris Vessey, Venkataraman Ramesh, and Robert L Glass. Research in information systems: An empirical study of diversity in the discipline and its journals. *Journal of management information systems*, 19(2):129–174, 2002.

[164] Judy Wajcman. Reflections on gender and technology studies: In what state is the art? *Social studies of science*, 30(3):447–464, 2000.

[165] Larry D Wall. Some financial regulatory implications of artificial intelligence. *Journal of Economics and Business*, 100:55–63, 2018.

[166] Ruotong Wang, F Maxwell Harper, and Haiyi Zhu. Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2020.

[167] Zikai Alex Wen, Jingyu Jia, Hongyang Yan, Yaxing Yao, Zheli Liu, and Changyu Dong. The influence of explanation designs on user understanding differential privacy and making data-sharing decision. *Information Sciences*, 2023.

[168] Marilyn Domas White, Emily E Marsh, Emily E Marsh, and Marilyn Domas White. Content analysis: A flexible methodology. *Library trends*, 55(1):22–45, 2006.

[169] C. Willig. Constructivism and 'the real world': Can they co-exist? *QMiP Bulletin*, (21), May 2016. This is a pre-publication version of the following article: Willig, C. (2016) Constructivism and 'The Real World': Can they co-exist?. QMiP Bulletin, 21.

[170] Stu Winby and Susan Albers Mohrman. Digital sociotechnical system design. *The Journal of Applied Behavioral Science*, 54(4):399–423, 2018.

[171] Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R O'Brien, Thomas Steinke, and Salil Vadhan. Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.*, 21:209, 2018.

[172] Christopher Woolard. The woolard review - a review of change and innovation in the unsecured credit market, Feb 2021.

[173] Tao Xiang, Yang Li, Xiaoguo Li, Shigang Zhong, and Shui Yu. Collaborative ensemble learning under differential privacy. In *Web Intelligence*, volume 16, pages 73–87. IOS Press, 2018.

[174] Jing Jian Xiao, Chuanyi Tang, Joyce Serido, and Soyeon Shim. Antecedents and consequences of risky credit behavior among college students: Application and extension of the theory of planned behavior. *Journal of Public Policy & Marketing*, 30(2):239–245, 2011.

[175] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. Towards effective differential privacy communication for users' data sharing

decision and comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 392–410. IEEE, 2020.

[176] Depeng Xu, Wei Du, and Xintao Wu. Removing disparate impact of differentially private stochastic gradient descent on model accuracy. *arXiv preprint arXiv:2003.03699*, 2020.

[177] Depeng Xu, Shuhan Yuan, and Xintao Wu. Achieving Differential Privacy and Fairness in Logistic Regression. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 594–599, San Francisco USA, May 2019. ACM.

[178] Elīna Zelčāne and Anita Pipere. Finding a path in a methodological jungle: a qualitative research of resilience. *International Journal of Qualitative Studies on Health and Well-being*, 18(1):2164948, 2023.

[179] Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, and Marianne Winslett. Functional mechanism: regression analysis under differential privacy. *arXiv preprint arXiv:1208.0219*, 2012.

[180] Lingchen Zhao, Lihao Ni, Shengshan Hu, Yaniiao Chen, Pan Zhou, Fu Xiao, and Libing Wu. Inprivate digging: Enabling tree-based distributed data mining with differential privacy. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 2087–2095. IEEE, 2018.

[181] T. Zhu, D. Ye, W. Wang, W. Zhou, and P. Yu. More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2020. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[182] Shoshana Zuboff. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, 2019.

Appendices

Appendix A

User Study

A.1 Information and Consent Form

Attitudes and Experiences with Loan Applications: UK context

School of Computer Science Ethics Reference: CS

Funded by: EPSRC

What is this research about?
This research aims to understand how people feel about loan applications, data sharing in this context, and how well they understand the decision process behind these decisions. To do this, a series of semi-structured interviews followed up with an online survey will be administered.

Participation in the research is voluntary.

Your participation may help us understand how people feel about loan applications and how the decision process is perceived. This is an exploratory study which will influence the rest of my research on evaluating the impact of implementation of Privacy Enhancing Technologies in the Consumer Credit Industry.

What will my participation involve?
If you chose to take part in the study you will

- Be involved in an interview, which will be held online over Microsoft Teams at a time of your convenience. The interview should take around 30~45 minutes. The interview will ask about your previous experiences applying for loans.
- After the interview you will be emailed a survey which should take 10~15 minutes to gather some quantitative data about the usage of different data sources in loan applications.

A 15£ voucher will be gifted as reimbursement for your time after both activities are completed.

You may withdraw from the study at any time and do not have to give reasons for why you no longer want to take part. If you wish to withdraw please contact the researcher who gathered the data. If you receive no response from the researcher please contact the School of Computer Science's Ethics Committee.

If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

How will the data be used?
The results of the research will be disseminated via conference presentations and journal publications. Your data may be archived and reused in future for purposes that are in the public interest, or for historical, scientific or statistical purposes. The data will be stored on password protected University of Nottingham servers.

Privacy Notice

The University of Nottingham is committed to protecting your personal data and informing you of your rights in relation to that data. The University will process your personal data in accordance with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018 and this privacy notice is issued in accordance with GDPR Articles 13 and 14.

The University of Nottingham, University Park, Nottingham, NG7 2RD is registered as a Data Controller under the Data Protection Act 1998 (registration No. Z5654762, https://ico.org.uk/ESDWebPages/Entry/Z5654762).

The University has appointed a Data Protection Officer (DPO). The DPO's postal address is:

Data Protection Officer,
Legal Services
A5, Trent Building,
University of Nottingham,
University Park,
Nottingham
NG7 2RD

The DPO can be emailed at dpo@nottingham.ac.uk

Why we collect your personal data. We collect personal data under the terms of the University's Royal Charter in our capacity as a teaching and research body to advance education and learning. Specific purposes for data collection on this occasion are for a research project on the personal understanding of data.

The legal basis for processing your personal data under GDPR. Under the General Data Protection Regulation, the University must establish a legal basis for processing your personal data and communicate this to you. The legal basis for processing your personal data on this occasion is Article 6(1e) processing is necessary for the performance of a task carried out in the public interest.

How long we keep your data. The University may store your data for up to 25 years and for a period of no less than 7 years after the research project finishes. The researchers who gathered or processed the data may also store the data indefinitely and reuse it in future research.

Who we share your data with. Extracts of your data may be disclosed in published works that are posted online for use by the scientific community. Your data may also be stored indefinitely by members of the researcher team and/or be stored on external data repositories (e.g., the UK Data Archive) and be further processed for archiving purposes in the public interest, or for historical, scientific or statistical purposes.

How we keep your data safe. We keep your data securely and put measures in place to safeguard it. These safeguards include anonymization of data and encryption of devices on which your data is stored.

Your rights as a data subject. GDPR provides you, as a data subject, with a number of rights in relation to your personal data. Subject to some exemptions, you have the right to:

- withdraw your consent at any time where that is the legal basis of our processing, and in such circumstances you are not obliged to provide personal data for our research.
- object to automated decision-making, to contest the decision, and to obtain human intervention from the controller.
- access (i.e., receive a copy of) your personal data that we are processing together with information about the purposes of processing, the

categories of personal data concerned, recipients/categories of recipient, retention periods, safeguards for any overseas transfers, and information about your rights.

- have inaccuracies in the personal data that we hold about you rectified and, depending on the purposes for which your data is processed, to have personal incomplete data completed
- be forgotten, i.e., to have your personal data erased where it is no longer needed, you withdraw consent and there is no other legal basis for processing your personal data, or you object to the processing and there is no overriding legitimate ground for that processing.
- in certain circumstances, request that the processing of your personal data be restricted, e.g., pending verification where you are contesting its accuracy or you have objected to the processing.
- obtain a copy of your personal data which you have provided to the University in a structured, commonly used electronic form (portability), and to object to certain processing activities such as processing based on the University's or someone else's legitimate interests, processing in the public interest or for direct marketing purposes. In the case of objections based on the latter, the University is obliged to cease processing.
- complain to the Information Commissioner's Office about the way we process your personal data.
-

If you require advice on exercising any of the above rights, please contact the University's data protection team: data-protection@nottingham.ac.uk

I have read in full the information sheet

Signature:_____

Consent

Taking part in the study

1. I have read and understood the project information sheet or it has been read to me.
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.
3. I understand that taking part in the study requires me to provide data and that this will involve completing an online questionnaire.

Use of my data in the study

1. I understand that data which can identify me will not be shared beyond the project team.
2. I agree that the data provided by me may be used for the following purposes:
    1. Presentation and discussion of the project and its results in research activities (e.g. project meetings, conferences).
    2. Publications and reports describing the project and its results
    3. Dissemination of the project and its results, including publication of data on web pages and databases.
3. I give permission for my words to be quoted for the purposes described above.

Reuse of my data

1. I give permission for the data that I provide to be reused for the sole purposes of future research and learning.
2. I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers

Security of my data

1. I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use.
2. I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me.

3. I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data.

Copyright
I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented.

Researcher's contact details
Name: Ana Rita Pena
Phone: 07599492130
Email: ana.pena@nottingham.ac.uk

I confirm that I have read the previous information and agree to take part in this study (tick the response below):

Yes-
No-

Signature: _____

A.2 Interview Guide

User Interviews Guide

Hello. My name is Ana Rita Pena and I'm a PhD student at The University of Nottingham. My PhD is about Privacy and Loan Applications, in particular I plan to evaluate the impact of implementing technologies to protect privacy in automated loan decisions.

We want to hear from people who have applied for loans in the UK in the past, to have a better understanding of how they feel about them and understanding of the decision process to try and bridge the gap between technical terms and the general user's understanding.

This interview is confidential and no one will be able to be identified when the research is written up. If you feel uncomfortable with any questions you do not have to answer it. We can stop the interview at any point. There is a £15 voucher for taking part. Interview will take approx. 45 mins.

Are you happy to go ahead?

Demographics - tax bracket you are inserted in,
gender,
ethnicity
education level

Tax band	Taxable income	Rate
Personal allowance	up to £12,500	0%
Basic rate	£12,501 to £50,000	20%
Higher rate	£50,001 to £150,000	40%
Additional rate	over £150,000	45%

1- INITIAL QUESTIONS

1.1- Have you ever applied for a loan in the UK?
1.1.1- Tell me about that experience.
1.1.2- [Why did you decide to take that loan? What was it for?]
1.1.3- [How important was it that loan was approved]
1.1.4- [What were the good and bad bits of applying for the loan? And why?]
1.1.5- [What impact did that have on you]
1.1.6- [Overall how satisfied where you with the experience]
1.1.7- [Is there anything you wish they would have done differently? and what would that mean for you?]
1.1.8- [Was this the only experience applying for a loan?]

1.2- From your understanding what makes a loan application be accepted or rejected? [aka What is the decision process like?]
1.2.1- [How do you feel about this?]

1.3- Do you feel like you were discriminated against?
1.3.1- [What is discrimination to you in this context]
1.3.2- My explanation of discrimination. - treating a person unfairly/differently because of who they are, or because they possess certain characteristics.
1.3.3- [Have you ever experienced it?]

SET TERMS USED IN THE INTERVIEW FROM THERE ONWARDS
CREDIT REPORT: A record of information such as your bill-paying history, length of your account with a company, any outstanding debt you have, any unpaid debts that have been registered with a court, any public record of having been sued, gone bankrupt, or failing to pay taxes (tax lien), and history of debt collection against you; it is used to determine your credit score.

CREDIT SCORE: A risk assessment tool used primarily by lenders. A high credit score means you are a low risk, are likely to get a loan or other service, and will have lower interest rates or more flexible terms of repayment. A low credit score could result in not getting a loan or other services, or paying more for such services. In other words, people with
"bad credit" may be charged a higher interest rate for any kind of loan, or denied a loan altogether.

ALGORITHMS: A sequence of steps for solving a problem (e.g., like a recipe); in digital terms, a set of computational or mathematical formulas that use data as their main ingredient, transforming these data (input) into desired outputs.

Automatic decision making – when a machine, for example computer, makes decisions based on rules (either defined by humans or defined by themselves) – NOT CLEAR ENOUGH

-Decisions made without the input of a human

DATA: Facts, details, statistics, or any information collected together for reference or analysis. Information

IMPLICIT BIAS: The unconscious attitudes, stereotypes, and unintentional actions (positive or negative) towards members of a group merely because of their membership in that group. These associations develop over the course of a lifetime beginning at a very early age through exposure to direct and
indirect messages.

BIASED DATA: When data is biased, we mean that the sample is not representative of the entire population. Can also just be the outcome of historical data.

**PRIVACY (ALSO KNOWN AS
DATA PRIVACY):** A human right that respects the right of people, including their data, to be left alone or kept to themselves. Privacy is also considered

to be culturally and historically defined, meaning that data sharing practices might be considered perfectly OK for one group but not at all appropriate for another.

Machine Learning gives computer systems the ability to "learn" (i.e. progressively improve performance on a specific task) with data, without being explicitly programmed.
It uses algorithms that can learn from and make predictions on data to make decisions that are not simply the result of following instructions.

Artificial Intelligence (AI) is the goal of creating computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. The automation of natural intelligences (such as that of humans) remains highly contentious, and the scope of AI is much disputed.

Discrimination means treating a person unfairly because of who they are, or because they possess certain characteristics. If you have been treated differently from other people only because of who you are or because you possess certain characteristics, you may have been discriminated against.

Discrimination can occur in different forms:

- direct discrimination
- indirect discrimination
- discrimination by association
- discrimination by perception
- harassment
- victimisation

for discrimination in the context of algorithmic discrimination relate to a-levels example

Data is a general term that means pieces of information. It can be the content you post on social media, your account information, your IP address etc. A data record is the collection of different types of data of a individual, usually associated with their finances. Banks and credit card companies use the information in the credit record to give an applicant a credit score (a number which represents the probability that said person will repay the loan). It is based on the credit score that a loajn application is accepted.

2 --- MORE AND MORE A LOT OF TASKS IN A WIDE RANGE OF JOBS AND INDUSTRYS HAVE BEEN AUTOMATED. THIS INCLUDES THE FINANCIAL SECTOR. I WILL NOW FOCUS ON THE AUTOMATION OF CREDIT LOAN APPLICATIONS AND DECISIONS. ---

2.1- [To what extent are you aware of the automation of loan application decisions?]
2.1.1.- How do you feel about loans applications decisions being automated?
2.1.2- [What is your understanding of this process ?] aka [Do you know how these decisions are made by computers?]

2.2- Which information have you shared previously when applying for a loan?
2.2.1- [How do you feel about sharing said data?]
2.2.2- [How do you feel about other types of data being used for the decision process?]

THERE ARE A LOT OF NEW DATA THAT HAS RECENTLY STARTED BEING USED OR MIGHT BE USED IN THE CLOSE FUTURE, FOR EXAMPLE THERE IS DIFFERENT TYPES OF FINANCIAL DATA (FROM PAYMENT HISTORIES OF UTILITIES, PHONE, COUNCIL TAX) TO TRANSACTIONAL SCORE AND CCJ(COUNTY COURT JUDGEMENT) SATISFACTIONS. HOWEVER NON-FINANCIAL DATA IS ALSO BEING USED FOR EXAMPLE INTERnET AND DEVICE INFORMTAION, SOCIAL MEDIA TO MORE MIXED DATA TYPES LIKE BEHAVIOURAL SCORE OR CONTEXTUAL INFORMATION ABOUT DEFAULT.

2.3- [How do you feel about these specific data types being used for loan applications. Do any of them stand out? and why?]

Financial data types	Mixed data types	Non-financial data types
Payment history: Phone and Broadband Social Housing/Private Rental Utilities Vehicle Finance Council Tax	Behaviour Score – Predictive score that can take into account financial information with possibly non-financial information e.g. change of address or marital status Delinquency- Contextual information about default, reason and length	Internet and device Omni Channel – phone, email, SMS Form completion and detailed internet and device data Mobile phone location Social Media
Income/Account info: Current Account Turnover (CATO)- estimates income based on income and expenditure HMRC – data about income based on UK's taxes		
Other: CCJ satisfactions- cases where the defendant has paid outstanding debt in full Transactional – behaviour predicted based on transactional data		

Note: Categories in colour have not yet been implemented. The colour ranks the probability of implementation. High, Medium, Low

4- FINAL SECTION

4.1- [Based on your experiences and the topics discussed in this interview do you feel like loan applications are fair?]
4.1.1- [Probe their definition of fairness]
4.1.2- [Do you think this would change if the decision was made by a human/computer?]- maybe don't ask

4.2- [Do you feel like you understand enough about how loan decisions are made? Either human or computer made]
4.2.1- [What do you need more information of?]

4.3- [What do you think the impact be if you had a better understanding of this decision process?]

Thank you so much for taking part in the interview.
Is there anything else you would like to add?

A.3 Post-Interview Survey

Attitudes and Experiences with Loan Applications

Information sheet

Attitudes and Experiences with Loan Applications: UK Context

School of Computer Science Ethics Reference: CS-XXXXXXX

Funded by: UK Engineering and Physical Sciences Research Council

For each question there will be a scenario, usually a task to be completed, and a choice of different algorithms.There are **20 questions** and the survey should take around **15 minutes to complete**.

The **aim** of this survey is to explore in detail how comfortable participants are with sharing different data types in the context of loan applications.This survey is a follow-up to a semi-structured interview and hence it should only be partaken by those who took part in the interview.

Participation in this survey is **voluntary** and there are **no forseable risks involved** in participation. The survey is aimed at any person **over the age of 18**.

All the **data is anonymised** and will be stored on password protected University of Nottingham servers.This means that you will not be able to be identified from the responses you provide in this survey.

The data will be stored in the JISC platform during the length of the study and later on a team that only the research team will have access to.

The results of the research will be disseminated via conference presentations and journal publications.Your data may be archived and reused in future for purposes that are

in the public interest, or for historical, scientific or statistical purposes.

Withdrawal:

It is **possible to withdraw from the survey at any time**. If you wish to withdrawal from the survey after submission please email the Researcher with your Individual unique code.

If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

Privacy Notice:
The University of Nottingham is committed to protecting your personal data and informing you of your rights in relation to that data. The University will process your personal data in accordance with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018 and this privacy notice is issued in accordance with GDPR Articles 13 and 14.

The University of Nottingham, University Park, Nottingham, NG7 2RD is registered as a Data Controller under the Data Protection Act 1998 (registration No. Z5654762, https://ico.org.uk/ESDWebPages/Entry/Z5654762).

The University has appointed a Data Protection Officer (DPO). The DPO's postal address is:

Data Protection Officer,
Legal Services
A5, Trent Building,
University of Nottingham,
University Park,
Nottingham
NG7 2RD

The DPO can be emailed at dpo@nottingham.ac.uk

Why we collect your personal data. We collect personal data under the terms of the University's Royal Charter in our capacity as a teaching and research body to advance

education and learning. Specific purposes for data collection on this occasion are for a research project on the personal understanding of data.

The legal basis for processing your personal data under GDPR. Under the General Data Protection Regulation, the University must establish a legal basis for processing your personal data and communicate this to you. The legal basis for processing your personal data on this occasion is Article 6(1e) processing is necessary for the performance of a task carried out in the public interest.

How long we keep your data. The University may store your data for up to 25 years and for a period of no less than 7 years after the research project finishes. The researchers who gathered or processed the data may also store the data indefinitely and reuse it in future research.

Who we share your data with. Extracts of your data may be disclosed in published works that are posted online for use by the scientific community. Your data may also be stored indefinitely by members of the researcher team and/or be stored on external data repositories (e.g., the UK Data Archive) and be further processed for archiving purposes in the public interest, or for historical, scientific or statistical purposes.

How we keep your data safe. We keep your data securely and put measures in place to safeguard it. These safeguards include anonymization of data and encryption of devices on which your data is stored.

Your rights as a data subject. GDPR provides you, as a data subject, with a number of rights in relation to your personal data. Subject to some exemptions, you have the right to:

- withdraw your consent at any time where that is the legal basis of our processing, and in such circumstances you are not obliged to provide personal data for our research.
- object to automated decision-making, to contest the decision, and to obtain human intervention from the controller.
- access (i.e., receive a copy of) your personal data that we are processing together with information about the purposes of processing, the categories of personal data concerned, recipients/categories of recipient, retention periods, safeguards for any overseas transfers, and information about your rights.
- have inaccuracies in the personal data that we hold about you rectified and, depending on the purposes for which your data is processed, to have personal incomplete data completed
- be forgotten, i.e., to have your personal data erased where it is no longer needed, you

withdraw consent and there is no other legal basis for processing your personal data, or you object to the processing and there is no overriding legitimate ground for that processing.
- in certain circumstances, request that the processing of your personal data be restricted, e.g., pending verification where you are contesting its accuracy or you have objected to the processing.
- obtain a copy of your personal data which you have provided to the University in a structured, commonly used electronic form (portability), and to object to certain processing activities such as processing based on the University's or someone else's legitimate interests, processing in the public interest or for direct marketing purposes. In the case of objections based on the latter, the University is obliged to cease processing.
- complain to the Information Commissioner's Office about the way we process your personal data.

If you require advice on exercising any of the above rights, please contact the University's data protection team: data-protection@nottingham.ac.uk

I have read in full the Information Sheet:

- ○ Yes
- ○ No

Consent

Consent:

Taking part in the study

1. I have read and understood the previous page information.
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time before submission, without having to give a reason.
3. I understand that taking part in the study requires me to provide data and that this will involve completing an online survey.

Use of my data in the study

1. I agree that the data provided by me may be used for the following purposes:
    1. Presentation and discussion of the project and its results in research activities (e.g. project meetings, conferences).
    2. Publications and reports describing the project and its results
    3. Dissemination of the project and its results, including publication of data on web pages and databases.

Reuse of my data

1. I give permission for the data that I provide to be reused for the sole purposes of future research and learning.
2. I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers

Security of my data

1. I understand that safeguards will be put in place to protect my data during the research, and if my data is kept for future use.
2. I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data.

Copyright

I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented.

Researcher's contact details

Name: Ana Pena

Phone: 07599492130

Email: ana.pena@nottingham.ac.uk

I have read the above document and:

- ○ Agree, and wish to procede with the survey
- ○ Do not agree, and will withdraw from the survey

Unique Individual Code

In order to be able to match your survey data with your interview please insert the Unique Individual Code given to you at the end of the Interview:

Recently added Data Sources

In this section I will ask you about data sources that have recently been used for laon applications in the UK (not all by all companies).

Do you think that it is acceptable to share the following types of data for use in loan application decisions?

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	N/A
Current Account Turnover (income and expenditure information used to estimate income)	○	○	○	○	○	○
Phone and Broadband payment history	○	○	○	○	○	○
Social Housing Data payment history	○	○	○	○	○	○
Private Rental payment history	○	○	○	○	○	○
Utilities payment history	○	○	○	○	○	○
Vehicle finance payment history	○	○	○	○	○	○
Behaviour score (predictive score based on credit usage pattern. Might include non-financial information)	○	○	○	○	○	○
Internet and Device (e.g. IP address, used to estimate location)	○	○	○	○	○	○
Omni Channel (e.g. phone, email, SMS)	○	○	○	○	○	○

Is there any combination of data sources that you WOULD NOT like to be used together? If so which? And why?

Is there any combination of data sources you WOULD like to be used together? If so which? And why?

Potential New Data Sources

In this section I will ask you about data sources that are not yet used in loan applications in the UK. Some have been used in other countries and some might come into use in the near future.

Do you think that it is acceptable to share the following types of data for use in loan application decisions?

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree	N/A
CCJ satisfactions (recording of County Court Judgement when debt has been paid in full)	○	○	○	○	○	○
Council Tax	○	○	○	○	○	○
HMRC data (income data held by UK's tax authority)	○	○	○	○	○	○
Transactional Behaviour (Behavioural prediction derived from user's transaction history)	○	○	○	○	○	○
Delinquency (contextual information, reason and length)	○	○	○	○	○	○
Form completion and detailed internet and device data	○	○	○	○	○	○
Mobile Phone Location	○	○	○	○	○	○
Social Media	○	○	○	○	○	○

Is there any combination of data sources that you WOULD NOT like to be used together? If so which? And why ?

Is there any combination of data sources you WOULD like to be used together? If so which? And why?

Final page

Thank you for submiting your answers.

Ana Pena is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (UKRI Grant No. EP/S023305/1).

If you wish to know more about the research done at the Horizon Centre for Doctoral Training: https://highlights.cdt.horizon.ac.uk/

Withdrawal:

If you wish to withdrawal form the study after submitting your answers email ana.pena@nottingham.ac.uk with your Unique Individual Code.

If you were brought here after the Privacy and Consent page no further action is required.

A.4 Equality Monitoring Form

Equality and diversity monitoring form

In this study we want to meet the aim of building an accurate picture of the study population encouraging equality and diversity.

The researcher (Ana Rita Pena, ana.pena@nottingham.ac.uk) needs your help and co-operation to enable it to do this, but filling in this form is voluntary (and you are free to not answer any questions you don't feel comfortable with).

The form will not be used for the analysis of the study data. It will only be used to monitor the equality and diversity of the study population.

Please return the completed form by email to ana.pena@nottingham.ac.uk (you do not need to include your Unique Identifiable Code).

Gender Man ☐ Woman ☐ Intersex ☐ Non-binary ☐ Prefer not to say ☐ If you prefer to use your own term, please specify here ……

Are you married or in a civil partnership? Yes ☐ No ☐ Prefer not to say ☐

Age 16-24 ☐ 25-29 ☐ 30-34 ☐ 35-39 ☐ 40-44 ☐ 45-49 ☐
50-54 ☐ 55-59 ☐ 60-64 ☐ 65+ ☐ Prefer not to say ☐

What is your ethnicity?
Ethnic origin is not about nationality, place of birth or citizenship. It is about the group to which you perceive you belong. Please tick the appropriate box

White

English ☐ Welsh ☐ Scottish ☐ Northern Irish ☐ Irish ☐
British ☐ Gypsy or Irish Traveller ☐ Prefer not to say ☐
Any other white background, please write in:

Mixed/multiple ethnic groups

White and Black Caribbean ☐ White and Black African ☐ White and Asian ☐
Prefer not to say ☐ Any other mixed background, please write in:

Asian/Asian British

Indian ☐ Pakistani ☐ Bangladeshi ☐ Chinese ☐ Prefer not to say ☐
Any other Asian background, please write in:

Black/ African/ Caribbean/ Black British

African ☐ Caribbean ☐ Prefer not to say ☐
Any other Black/African/Caribbean background, please write in:

Other ethnic group

Arab ☐ Prefer not to say ☐ Any other ethnic group, please write in:

Do you consider yourself to have a disability or health condition?

Yes ☐ No ☐ Prefer not to say ☐

What is the effect or impact of your disability or health condition on your ability to give your best at work? Please write in here:

The information in this form is for monitoring purposes only.

What is your sexual orientation?

Heterosexual ☐ Gay ☐ Lesbian ☐ Bisexual ☐

Prefer not to say ☐ If you prefer to use your own term, please specify here .....

What is your current working pattern?

Full-time ☐ Part-time ☐ Prefer not to say ☐

What is your flexible working arrangement?

None ☐ Flexi-time ☐ Staggered hours ☐ Term-time hours ☐

Annualised hours ☐ Job-share ☐ Flexible shifts ☐ Compressed hours ☐

Homeworking ☐ Prefer not to say ☐ If other, please write in:

Do you have caring responsibilities? If yes, please tick all that apply

None ☐ Primary carer of a child/children (under 18) ☐

Primary carer of disabled child/children ☐

Primary carer of disabled adult (18 and over) ☐ Primary carer of older person ☐

Secondary carer (another person carries out the main caring role) ☐

Prefer not to say ☐

Appendix B

Industry Study

B.1 Information Sheet

UK Consumer Credit Industry Stakeholder Consultation

School of Computer Science Ethics Reference: CS

Funded by: EPSRC

What is this research about?

Research activity aimed at better understanding the Consumer Credit Industry, understand the relationship between the different Stakeholders as well as their views on Differential Privacy and its implementation.

This activity focuses on the following Stakeholders in specific: Credit Industry, its Regulators and other Charities and Consumer Advice institutions.

What will my participation involve?
If you chose to take part in the study you will:

- Be involved in an interview, which will be held online over Microsoft Teams at a time of your convenience. The interview should take around 60~90 minutes. The interview will ask about your experiences working in the Consumer Credit Industry (no need to discuss commercially sensitive information) and will ask about your opinions on Differential Privacy, a technology which will be explained during the interview.

The interview is divided into two parts: the initial part will ask about your role, day to day work and the workings of the Consumer Credit Industry more generally. The second part focuses more on Privacy and will ask about which actions are taken to promote consumer's privacy, and I will explain what Differential Privacy is and a couple of scenarios to get your thoughts on its possible implementation in the industry.

Participation in the research is voluntary. Participation in the study needs to be approved by participant's line manager.

Your participation may help us understand the workings of the Consumer Credit Industry as well as potential impacts of the implementation of Differential Privacy.

This is a qualitative study which will influence the rest of my research on evaluating the impact of implementation of Privacy Enhancing Technologies in the Consumer Credit Industry.

A 15£ voucher will be gifted as reimbursement for your time or alternatively this amount will be donated to the Citizen's Advice Charity.

You may withdraw from the study at any time and do not have to give reasons for why you no longer want to take part. If you wish to withdraw please contact the researcher who gathered the data. If you receive no response from the researcher please contact the School of Computer Science's Ethics Committee.

If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

How will the data be used?
The results of the research will be disseminated via conference presentations and journal publications. Your data may be archived and reused in future for purposes that are in the public interest, or for historical, scientific or statistical purposes. The data will be stored on password protected University of Nottingham servers.

B.2 Consent Form

CONSENT FORM

University of Nottingham
UK | CHINA | MALAYSIA

Date: 24/03/2022

Project: UK Consumer Credit Industry Stakeholder Consultation

School of Computer Science Ethics Reference: CS-2021-R3

Funded by: EPSRC

Please tick the appropriate boxes	**Yes**	**No**
1. Taking part in the study		
a) I have read and understood the project information sheet dated 24/03/2022 or it has been read to me. I have been able to ask questions about the study and my questions have been answered satisfactorily.	☐	☐
b) I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
c) I understand that taking part in the study requires me to provide data and that this will involve a 60~90 minute interview	☐	☐
2. Use of my data in the study		
a) I understand that data which can identify me will not be shared beyond the project team.	☐	☐
b) I agree that the data provided by me may be used for the following purposes:		
– Presentation and discussion of the project and its results in research activities (e.g., in supervision sessions, project meetings, conferences).	☐	☐
– Publications and reports describing the project and its results.	☐	☐
– Dissemination of the project and its results, including publication of data on web pages and databases.	☐	☐
c) I give permission for my words to be quoted for the purposes described above.	☐	☐

Please tick the appropriate boxes	Yes	No
3. Reuse of my data		
a) I give permission for the data that I provide to be reused for the sole purposes of future research and learning.	☐	☐
b) I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers	☐	☐
4. Security of my data		
a) I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use.	☐	☐
b) I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me.	☐	☐
c) I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data.	☐	☐
5. Copyright		
a) I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented.	☐	☐

6. Signatures (sign as appropriate)

Name of participant (IN CAPITALS)	Signature	Dat

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Name of researcher (IN CAPITALS)	Signature	Date

7. Researcher's contact details

Name: Ana Rita Pena

Phone: 07599492130

Email: ana.pena@nottingham.ac.uk

B.3 Interview Guide

Stakeholder Consultation Study

Introduction

Hello. My name is Ana Rita Pena and I'm a PhD student at The University of Nottingham. My PhD is about Privacy and Loan Applications, in particular I plan to evaluate the impact of implementing technologies to protect privacy in automated loan decisions.

We want to hear from people who work or have worked within and around the Consumer Credit Industry. This study has two main aims, the first is to have a better understanding of the inner workings and interactions between different stakeholders in the Industry and the second is to understand the role of Privacy in the industry, and in specific attitudes on the implementation of Differential Privacy.

This interview is confidential, and no one will be able to be identified when the research is written up. If you feel uncomfortable with any questions you do not have to answer it. We can stop the interview at any point. The participation in this study requires the approval of you line manager. You are not expected to give away any commercially sensitive information.
There is a £15 voucher for taking part, or that value can be donated to the Citizens Advice Charity. The interview will take between 60 and 90 mins.
Are you happy to go ahead?

Part I- General

- Could you please describe your institution's role and general working within the financial sector? [Follow up specific credit]
- Could you describe your role and day to day work ?

- What are your thoughts on the use of technology (or AI/ML in specific) in the banking and consumer credit sector? [pros and cons if not mentioned]
- [Do they align with company thoughts? Or general industry?]

- How do you perceive your institution's (and/or stakeholders) understanding of new technology used in the sector ?
  [Are you up to date with the state of the art algorithms implemented in the Industry?]

- How much do you know about which algorithms/tech are used in practise in the sector ? [come up with eg.. of what I'm talking about (To do risk scoring ML,LR), or alternative ways of questioning]
- [Or are you aware of what technologies your competitors use?]

- Where in the tech development stage do regulators and regulation come into play?

- How would you describe the relationship between your institution and the Financial Industry?

- In your perspective what is the purpose of the Consumer Credit Industry? [Does it accomplish it?/what would it ideally accomplish?]
- [similar question regarding purpose but from institution perspective on above questions]
- Do you think all Stakeholders see the purpose of the consumer credit industry in the same way ? [and what are the consequences of this?]

Part II – Privacy

- In which way is privacy taken into account in your institution ? [Is it a topic that is commonly talked and discussed ?] [Does your institution take any proactive privacy measures which are not required by regulators and law ?]

- Are you aware of which Privacy Enhancing Technologies are implemented in your Institution? Could you share your understanding of them?

Differential Privacy Communication with Stakeholders

- Introduction to the Concept and Privacy-Accuracy Trade-off

Privacy has different meanings in different contexts, in the digital one it usually means not being able to be identified out of a big group of people. For example, think of an online shop that has all the information from its costumers' purchases if someone saw that data and was able to identify a specific costumer most consumers would say that their privacy was breached.

Due to several controversies (we will see one of them in the next slide), there were several techniques to enhance privacy that have been developed. One of them is Differential Privacy.

However, most privacy methods cannot prevent either reconstruction attacks (combinations of different public datasets to identify individuals in them) or model inversion attacks (being to gather information on training data points by having access to the model and a secondary dataset), in these cases anonymisation is not enough.

One of the most famous reconstruction attacks was able to combine anonymised medical data that included performed medical procedures, prescribed medications, ethnicity, and people's gender, date of birth, and ZIP code. And a voter's registration list with demographics to get information on the Governor Weld's medical information.

In the context of Consumer Credit implementing Differential Privacy could mean that if companies were required to share their model with either regulators or the public its consumer's privacy would still be protected, for example.
Now we will discuss more how differential privacy works and prevents these types of attacks.

A good way to understand the way these technologies work is by thinking of looking through fogged glass. Let's look at a group of people through some glass. When the glass is perfectly clear it is easy to differentiate between different people and identify them even those that look more similar. When the glass starts fogging up it becomes hard to

distinguish between very similar people. The more fog there is on the glass the harder it is to identify specific people even if they don't look that similar, in the end everyone would look the same, just like a stain.

This is the way that differential privacy works, it adds specifically designed noise (the fog) so that very similar things are hard to distinguish and hence protecting the privacy of individuals. The amount of noise/fog added can be chosen according to the balance of accuracy (being able to differentiate people well) and privacy level we wish to maintain. I'll explain a bit more what this "noise" looks like when we add some more technical detail.

Any questions so far?

- Technical Intro details

Now that we have a general intuition of how Differential Privacy works, we'll add a bit more technical detail and start to understand how does this work within a Machine Learning Context.

The most common use of Machine learning applications is in classification problems, e.g. is this an image of a dog?; whoever there are also different tasks like e.g. what is the probability that this person will repay a loan?

The way the computer answers these questions in most cases is by being shown pictures of dogs where they are told it's a dog and pictures of other objects, like cats for example, where they are told it is a cat.

As the computer does not have an abstract idea of what dogs or cats are what it does when its learning is looking for patterns and similarities between pictures that it is told belong to the same group.
This generalization of patterns and similarities happens in the form of a function (or a mapping) which collates information from all the different datapoints. When the computer is fed a datapoint it has not seen before it will input it to the function (after training) and its outcome will be our answer, e.g. yes it is a dog/not it is not a dog.

Questions so far?

In the Machine Learning context, saying an algorithm is Differentially Private is a guarantee that if we have two different but very similar inputs, the outputs of these through the function will be close together. And what this means in terms of privacy is that simply by looking at the outcomes (or labels) we will not be able to distinguish with certainty which input it came from. And this applies for every possible pair of inputs that are similar.

This goal is achieved by adding noise (our fog in this context) to either the input, the function, or the output of our computer algorithm. So if we consider the output for examples, let's say our function without considering privacy gives us an answer of 12 (if we were predicting someone's age based on a picture), the private function will add a small number (noise) to the true answer so it private answer would be 11 or 13 for example.

How much noise we had depends on several factors, but in general the more noise, the more privacy and the less accuracy (as described in fogged glasses examples).

Accuracy here means the percentage of time that our computer gets the answer right. The decrease of accuracy within the Consumer Credit history can lead to providing loans to consumers that cannot afford them (which could lead to default and negative impact on the consumers financial life as well as a loss of money or lack of profit to the bank). Alternatively, if the decrease in accuracy meant that consumers that were originally falsely labelled as not being able to afford a loan are now granted that will be beneficial for both the lending agencies (more profit) and the consumer.

Questions:

- What are your thoughts on this technology from what you've heard so far?
- Could you see this technology being implemented in the Consumer Credit Industry? [why?, in which way?]

- Different accuracy for Subgroups

We have discussed the general trade-off between privacy and accuracy, and now we will talk a bit more about the way in which this accuracy drop can be distributed within different subgroups.

I will now briefly describe three different cases and then we will discuss their implications.

In an ideal scenario the accuracy drop from making a model private would be the same or proportional across different subgroups of our dataset, however there have been studies which report that is not always the case (in fact it does not tend to be).

Scenario 1: Different subgroups in our training and testing datasets have different levels of accuracy, this can be due to some of the subgroups having a smaller sample size. When we create a private version of this model, the subgroups which had lower accuracy to start off with have a bigger decrease in accuracy than the others.

Scenario 2: Different subgroups in our training and testing datasets have different levels of accuracy, this can be due to some of the subgroups having a smaller sample size. When we create a private version of this model, the subgroups which had lower accuracy to start off with have a smaller decrease in accuracy than the groups who had higher accuracy to start.

Scenario 3: Different subgroups in our training and testing datasets have different levels of accuracy. When we create a private version of this model, there is no general trend which explains the way the accuracy drop distribution.

For all the scenarios we can adjust how much privacy we require. When we require a lot of privacy the lines between the private case and non-private case will be very separated and as we decrease the privacy requirements the lines will become closer together.

There are a variety of factors which can affect the distribution of the accuracy drop from the training data, including the different subgroups sizes to the specific algorithm that is implemented and the way it adds noise among others.
Current research in DP aims to better the accuracy-privacy trade-off for different types of models, as well as better access and mitigate the disparate accuracy drops.

Questions:

- Who do you think would benefit from this technology being implemented? who would be detrimented in all scenarios?
- Could you still see DP being implemented in Industry/your institution?

Stakeholder Pros and Cons here

- How would the working of your institution change to accommodate DP?
- How do you think the implementation of DP would impact the whole Consumer Credit Ecosystem?
- How do you think the implementation of DP could impact the consumer?
- Overall what are your thoughts on the positives and negative of Differential Privacy

B.4 DP Presentation Aid

UK CONSUMER CREDIT INDUSTRY STAKEHOLDER CONSULTATION

Ana Rita Pena

PhD Candidate at the
Horizon Centre for
Doctoral Training

ana.pena@nottingham.ac.uk

UKRI Engineering and Physical Sciences Research Council

horizon CENTRE FOR DOCTORAL TRAINING

University of Nottingham
UK | CHINA | MALAYSIA

DIFFERENTIAL PRIVACY

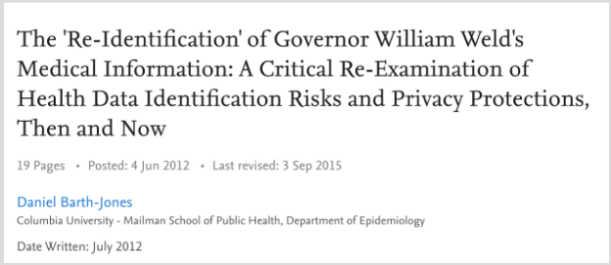
In this part of the interview we wil discuss a specific technology , Differential Privacy , which I study as part of my PhD.

WHY DIFFERENTIAL PRIVACY

In the digital context privacy usually means not being able to be identified

Several controversies led to the creation of different techniques to enhance privacy.

However most privacy methods cannot prevent either reconstruction attacks (example in the image) or model inversion attacks.



The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now

19 Pages • Posted: 4 Jun 2012 • Last revised: 3 Sep 2015

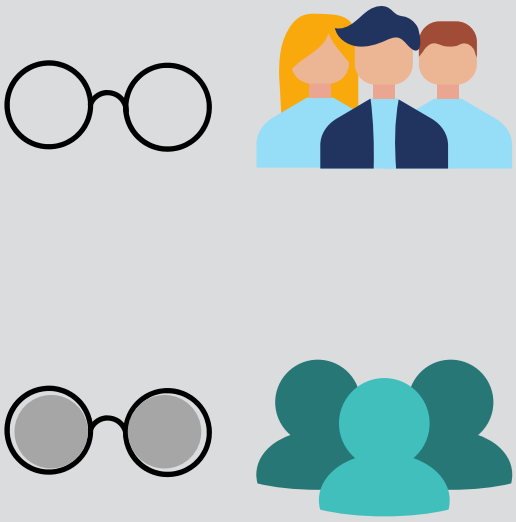
Daniel Barth-Jones
Columbia University - Mailman School of Public Health, Department of Epidemiology

Date Written: July 2012

CONCEPT INTRODUCTION

Differential Privacy works like looking through fogged glass.

The more fog there is on the glass the harder it is to identify and differentiate specific people out of a group.



TECHNICAL INTRODUCTION

Most common Machine Learning applications are classification problems (is this a dog or a cat?).

A computer answers this question by being shown pictures of both and finding patterns and similarities within the same group.

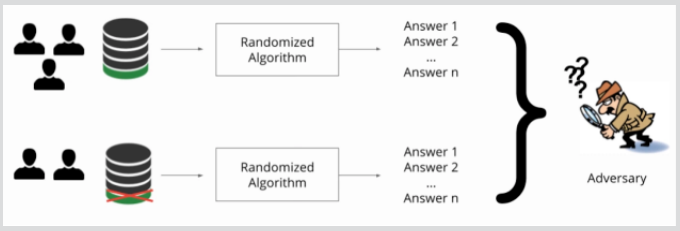
The generalisation of the patterns found happens in the form of a function.

When the computer is shown a picture it has not seen before , it puts it through the function and the outcome will be our answer (Yes, this is a dog).

TECHNICAL INTRODUCTION

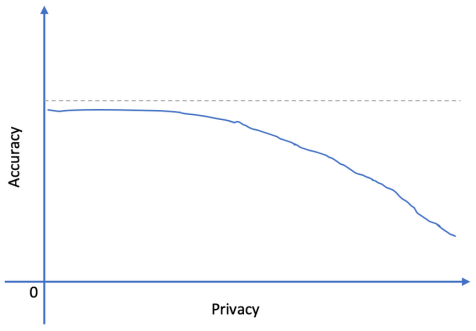
Saying an algorithm is Differentially Private is a guarantee that if we have two different but very similar inputs, the outputs of these through the function will be close together. And this applies for every possible pair of inputs that are similar.

The goal of privacy is achieved as simply by looking at the outcomes (or labels) we will not be able to distinguish with certainty which input it came from.



TECHNICAL INTRODUCTION

This goal is achieved by adding noise (our fog in this context) to either the input, the function, or the output of our computer algorithm. How much noise we had depends on several factors, but in general the more noise, the more privacy and the less accuracy (as described in fogged glasses examples).



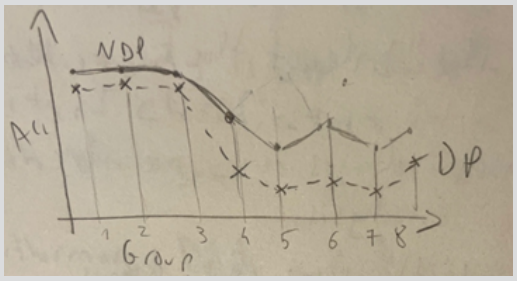

QUESTIONS

PRIVACY - ACCURACY TRADE-OFF

Now we will discuss and consider different scenarios where the accuracy drop stemming form the privacy implementation is differently distributed across different subgroups.

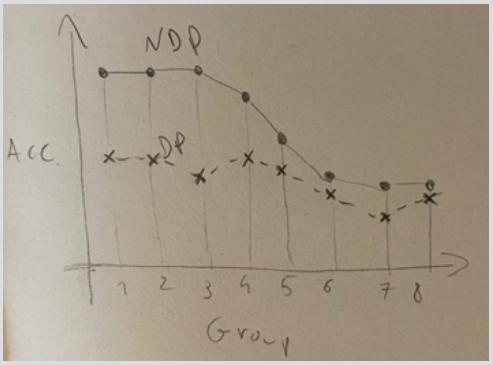
SCENARIO 1 - POOR BECOME POORER

Different subgroups in our training and testing datasets have different levels of accuracy, this can be due to some of the subgroups having a smaller sample size. When we create a private version of this model, the subgroups which had lower accuracy to start off with have a bigger decrease in accuracy than the others.



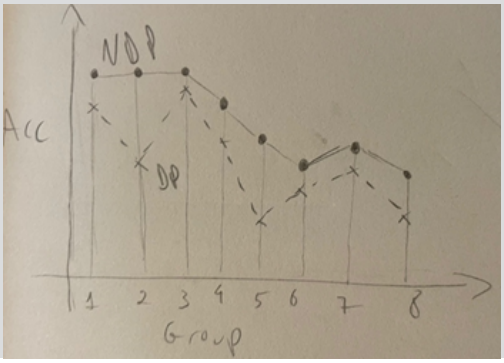
SCENARIO II - EQUALIZATION

When we create a private version of this model, the subgroups which had lower accuracy to start off with have a smaller decrease in accuracy than the groups who had higher accuracy to start.



SCENARIO III - RANDOM DISTRIBUTION

When we create a private version of this model, there is no general trend which explains the way the accuracy drop distribution.

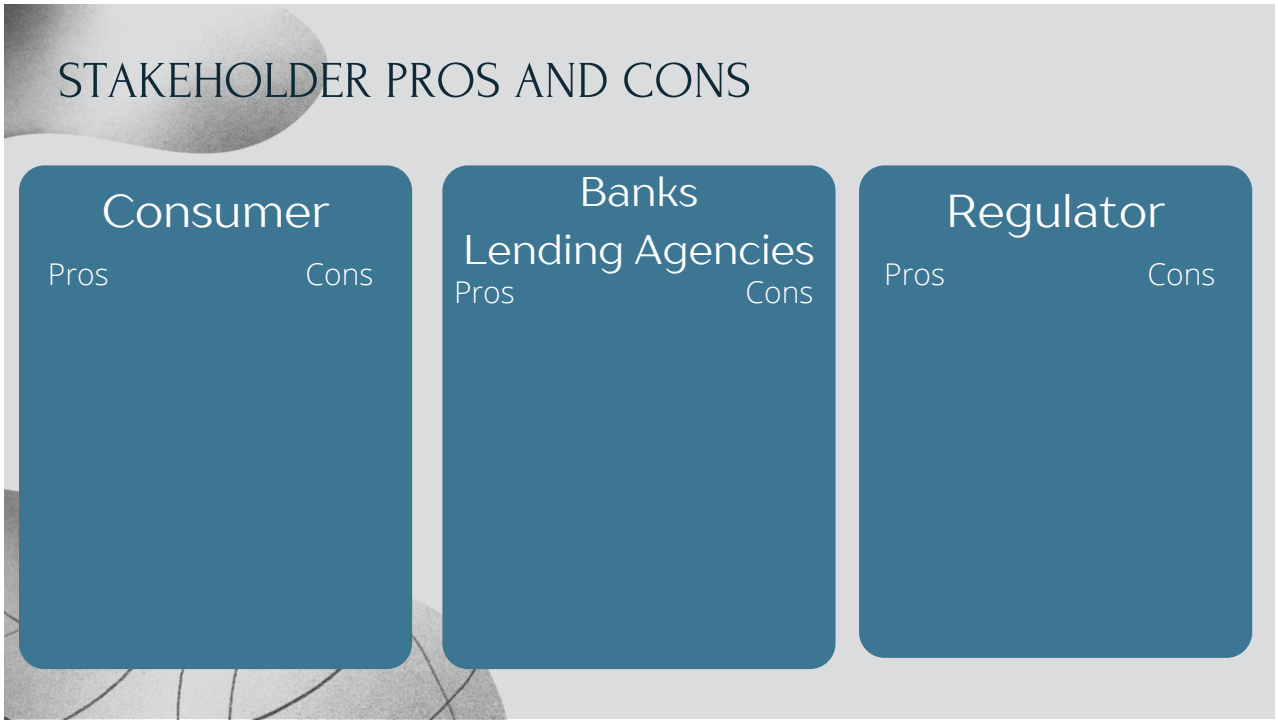


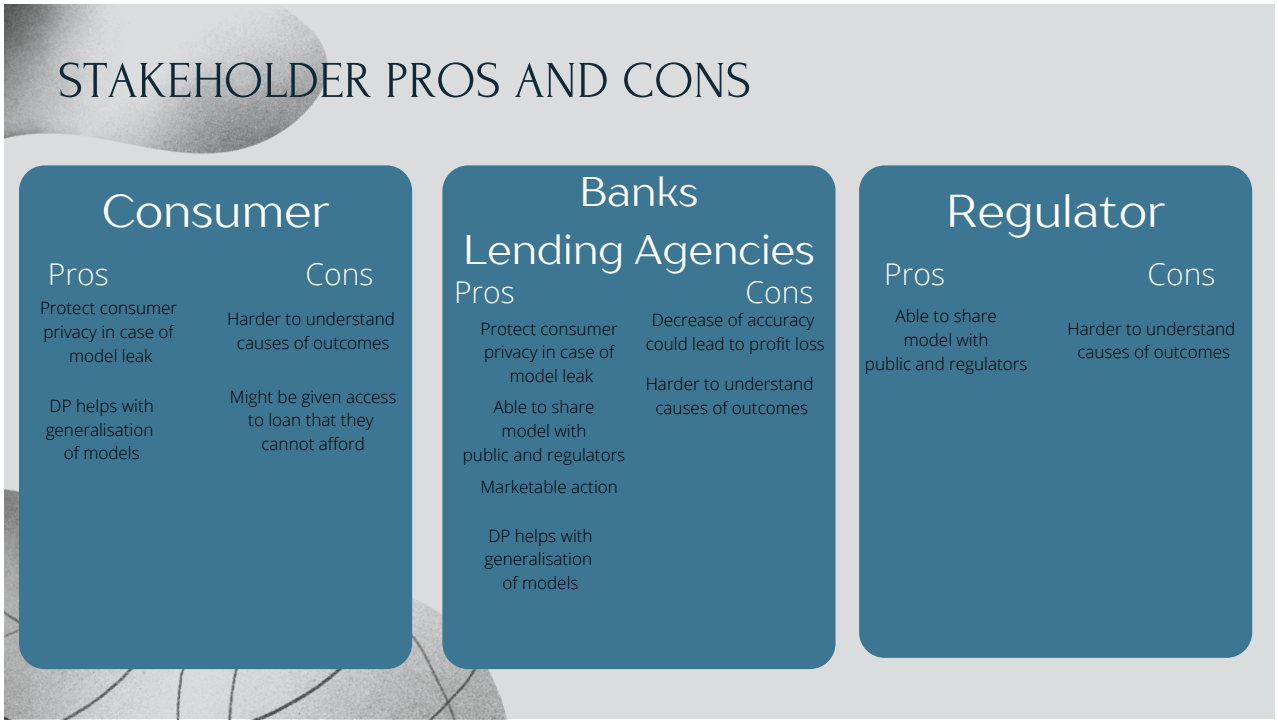
PRIVACY - ACCURACY TRADE-OFF

There are a variety of factors which can affect the distribution of the accuracy drop from the training data, including the different subgroups sizes to the specific algorithm that is implemented and the way it adds noise among others.

Current research in DP aims to better the accuracy-privacy trade-off for different types of models, as well as better access and mitigate the disparate accuracy drops.


QUESTIONS


STAKEHOLDER PROS AND CONS
Consumer
Pros
Cons
Banks
Lending Agencies
Pros
Cons
Regulator
Pros
Cons


STAKEHOLDER PROS AND CONS
Consumer
Pros
Protect consumer privacy in case of model leak
DP helps with generalisation of models
Cons
Harder to understand causes of outcomes
Might be given access to loan that they cannot afford
Banks
Lending Agencies
Pros
Protect consumer privacy in case of model leak
Able to share model with public and regulators
Marketable action
DP helps with generalisation of models
Cons
Decrease of accuracy could lead to profit loss
Harder to understand causes of outcomes
Regulator
Pros
Able to share model with public and regulators
Cons
Harder to understand causes of outcomes

QUESTIONS

Appendix C

Technical Study

C.1 Smooth Random Forest

The first algorithm implemented makes use of ϵ- differential privacy and smooth sensitivity [67]. As it is a Random Forest, the process of tree building does not need to query the data, only querying the leaf data for the majority class label making use of the Exponential Mechanism. Typically frequency queries such as label counts tend to be made differentially private by implementing the Laplacian mechanism, however the Smooth sensitivity of the Laplacian is the same as the global sensitivity (which is 1), instead for the Smooth Random Forest uses the Exponential Mechanism with utility function:

$$u(c, z) = \begin{cases} 1 & c = argmax_{i \in c} n_i \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

where z is a leaf of a decision tree and n_y is the number of datapoints with value c. This means that all values of c in the leaf node will be zero apart from the one with the most amount of datapoints, which will have a value of 1. The smooth sensitivity of this utility function is then:

$$S^*(u, z) = e^{-j\epsilon} \leq 1$$

where ϵ is the privacy budget of the query and j is the difference between the most frequent label and second most frequent.

By reducing or equating the smooth sensitivity (compared to the Laplacian mechanism) the amount of noise added in the leaf nodes query is reduced.

Each tree in the forest it built without needing to query the data, the building is stopped when the termination criteria is me, in this case it is the maximum depth. The maximum depth is automatically calculated based on the optimal depth based on the work of Fan et al. [63] which is extended in [67] to account for continuous covariates. Hence the optimal tree depth is:

$$d^* = \left(argmin_{d:n_s<s/2} s \left(\tfrac{s-1}{s}\right)^d\right) + \tfrac{r}{2}$$

where n_s is the expected number of continuous features (s) not tested, d is the tree's depth and r is the number of discrete covariates.

To make use of the parallel composition property of differential privacy, each tree is trained on a disjoint dataset, which then uses the full privacy budget allocated to the forest. In terms of optimal numbers of trees the authors have empirically tested different numbers and their combination with privacy budgets and found that between 1 and 10 tress the accuracy varies the most, and there tends to be a sweet spot between 30 and a 100

tress which changes to 100 to 300 trees with bugger privacy budget.

Empirically the Smooth Random Forest Algorithm outperformed a series of other models [94, 71] across a series of different datasets.

C.2 Gradient Boosting Decision Tree

There have been several different differentially private implementations of Gradient Boosting Decision Trees [107, 109, 173, 180], however [107] was chosen over the other models as it performs better. The GBDT algorithm achieves these results by implementing a novel boosting framework (Ensemble of Ensembles, seen on Figure C.1) and by introducing Gradient-based Data Filtering and Geometric Leaf Clipping to obtain closer bounds on the sensitivity of queries.

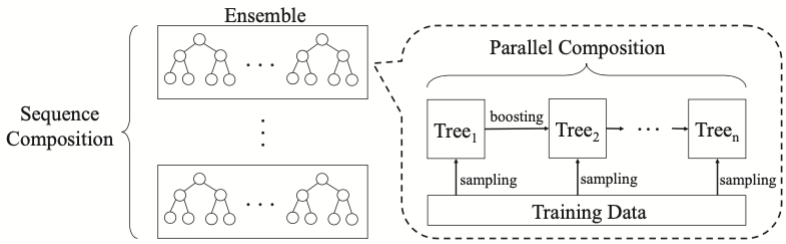


Figure C.1: Ensembles of Ensembles: Two level novel boosting framework used in GBDT. Figure reproduced from [107]

Non- private GBDT minimise a loss function with a regularizer term and have the gain as the splitting function, given in Definition 9.

Definition 9. (Gain of Split):

If I_L and I_R are the instances to the left and right nodes after a split and $I = I_L \cup I_R$ the gain of the split is:

$$G(I_L, I_R) = \frac{(\sum_{i \in I_L} g_i)^2}{\|I_L\| + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\|I_R\| + \lambda}$$

where g_i is the first order gradient statistics on the loss function and λ is the regularization parameter.

If the current nodes has achieved the tree's maximum depth or if slit gain is smaller than zero, tree building is stopped and it becomes a leaf node with an optimal value of:

$$V(I) = -\frac{\sum_{i \in I} g_i}{\|I\| + \lambda}$$

Algorithm 1: TrainSingleTree: Train a differentially private decision tree

Input: I: training data, $Depth_{max}$: maximum depth
Input: ε_t: privacy budget
$\varepsilon_{leaf} \leftarrow \frac{\varepsilon_t}{2}$ // privacy budget for leaf nodes
$\varepsilon_{nleaf} \leftarrow \frac{\varepsilon_t}{2Depth_{max}}$ // privacy budget for internal nodes
Perform gradient-based data filtering on dataset I.
for $depth = 1$ **to** $Depth_{max}$ **do**
  for *each node in current depth* **do**
    for *each split value* i **do**
      Compute gain G_i according to Equation (3).
      $P_i \leftarrow exp(\frac{\varepsilon_{nleaf} G_i}{2\Delta G})$
    /* Apply exponential mechanism */
    Choose a value s with probability $(P_s / \sum_i P_i)$.
    Split current node by feature value s.
for *each leaf node* i **do**
  Compute leaf value V_i according to Equation (4).
  Perform geometric leaf clipping on V_i.
  /* Apply Laplace mechanism */
  $V_i \leftarrow V_i + Lap(0, \Delta V / \varepsilon_{nleaf})$
Output: A ε_t-differentially private decision tree

Figure C.2: Single Tree building algorithm from the GBDT model.Figure reproduced from [107]

In individual tree construction in GBDT, summarised in Figure C.2, the initial step is Gradient-based Data Filtering which consists on filtering of the dataset to be used in training the specific tree. The filtering happens through a simple threshold value of the initial gradient value for each data point, where the threshold is the maximum possible 1-norm gradient in the initialisation and given by:

$$g_l^* = max_{k_p \in [-1,1]} \parallel \frac{\delta l(k_p,k)}{\delta k}|_{k=0} \parallel$$

where l is the loss function.

Following on from the data filtering the tree is built iteratively where the split are defined by the Exponential mechanism which uses the Gain of Split function as the utility function. Once tree building has stopped Geometric Leaf Clipping is performed in the leaf nodes. Based on a theoretical understanding of the optimal leaf values function, $V(I)$, the authors have found that leaf values in each tree approximately form a geometric sequence with common ration g_l^* and common ratio $(1-\eta)$ where η is a shrinkage rate. Geometric Leaf Clipping consists in replacing values larger than the threshold , which is $g_l^*(1-\eta)^{t-1}$ where t is the number of iteration, with the threshold value before applying the Laplacian Mechanism to the leaf value. AS GBDT are built sequentially should not influence the objective.

By performing Gradient-based Data Filtering and Geometric Leaf Clipping the sensitivity of both the Gain of Split and the Optimal Leaf Value had been reduced, which in turn reduces the amount of noise added in the Exponential and Laplacian mechanism when building a single tree. Half of the total privacy budget allocated for a single tree is used for the leaf nodes query, and the remaining half is equally split amongst the internal nodes, which uses a similar approach to [180, 115].

The Ensemble of Ensemble boosting framework is designed to both make use of the parallel and sequential composition properties of DP for budget allocation while still maintaining the effectiveness of boosting. Within each Ensemble T_e trees are trained on disjoint datasets in parallel, the ensembles, where N_e is the number of ensembles, are then built sequentially on the whole dataset as represented in Figure C.1. The privacy budget for each

tree is then ϵ/N_e , where ϵ is the total GBDT budget.

Appendix D

Focus Group Study

D.1 Information and Consent Sheet

University of Nottingham
UK | CHINA | MALAYSIA

School of Computer Science

University of Nottingham

Section B. Information to be provided to research participants

PROJECT TITLE: Differentially Private Consumer Credit Imaginaries

1. The research
a) Aims and objectives of the research
*The Differentially Private Consumer Credit Imaginaries Study consists of a focus group activity which **aims to understand how users/consumers perceive the implementation of Differential Privacy in different scenarios**, generated by a board game type of activity and following discussion.* *This research is the final study of my PhD, which focuses on Differential Privacy (a privacy enhancing technology) in the context of consumer credit.*
b) Funder information
This work was supported by the Engineering and Physical Research Council [Grant number EP/S023305/1]
c) Governance
This research has been approved by the School of Computer Science Research Ethics Committee (CS REC), ethics application ID ***insert ethics application ID once assigned***

2. Taking part in the research
The study will involve an in-person focus-group where a game board style activity will serve as a starting point to discuss participant's attitudes towards Differential Privacy in Credit (no background knowledge needed). *The study should take around 2 hours.* *The study will involve being audio and video recorded for analysis purposes.* *You will be remunerated for your time with a 25£ Amazon voucher*

3. Risks of participation
a) Risks

There is always a risk of unauthorised access to data.

b) Mitigation of risks

See section 5 for the measures we put in place to mitigate the risk of unauthorised access.

4. Purpose of data processing

a) Data collected

We collect the following categories of data during your participation in the research:

Focus group data – audio and video recording

b) Specific purposes for which the data are processed

Data collected during the research that identifies you may be:

- Analysed to meet the aims and objectives described in Section 1.
- Reviewed and discussed in supervision sessions between researchers and their supervisors or in research meetings between members of the research team, including project partners.
- If audio recordings are collected during the research, these may be transcribed and **anonymous** quotations of your spoken words may be used in scientific works, including presentations, reports and publications stored in databases and posted online, and in marketing materials that promote the research and its findings.
- If visual images that identify you are collected during the research, they may be used in scientific works, including presentations, reports and publications stored in databases and posted online, and in marketing materials that promote the research and its findings; you will **not** be named if visual data is used for these purposes and you may **opt out** in Section 9b.

c) Automated decision-making and profiling

NA

d) Legal basis for processing your data

We collect personal data under the terms of the University of Nottingham's Royal Charter and in our capacity as a teaching and research body to advance education and learning. We thus process your data on the legal basis that our research is in the public interest, we have legitimate interests and / or that you consent to data processing in freely and voluntarily participating in our research activities.

5. Storage and retention of your data

a) Data protection measures

We put the following organisational and / or technical safeguards in place to protect your data and your identity to the best of our ability:

i) All data stored digitally will be encrypted and password protected and all physical data will be stored in a secure location.

ii) *Describe any other organisational and/or technical safeguards that will be put in place to secure the participant's data and identity, e.g., forms stored in locked cupboards, anonymisation or pseudonymisation procedures, etc. (delete this text if not applicable).*

D.2 Interview Guide

Interview Guide

Hi everyone, my name is Ana Rita Pena, most people call me Rita and I am a PhD researcher at the University of Nottingham. My works focuses on evaluating the potential use of Differential Privacy (which is a technology used to protect people's privacy) in the Credit Industry (specifically consumer credit so for example credit cards and personal loans). I have spent the last three years looking into this topic and I have interviewed both Industry and consumers as well as doing some more technical work on the computer to understand this technology itself. In this study we are going to have a series of activities to talk about the lending industry in general, we will talk about the behind the scenes of an application process and then I will tell you a little bit about the technology (as I don't assume any prior knowledge) and then we will discuss it in a bit more detail as part of an activity.

Part 1 - Introduction

1. Can each of introduce yourselves and just let us know your name?
2. What are your thoughts on banks and lenders and the credit industry in general? (You can refer to your experiences but only if you wish to)
3. In a scale of 1 to 5 where 1 is not confident and 5 is very confident where do you position yourselves in relation to your knowledge of the loan application process? Why is this?

Part 2 - Game + Differential Privacy

Explanation of application process with board and explanation cards. (Power Point Aid)

Discussion of Differential Privacy (Power Point)

Explanation of the game

COMFORT BREAK

Game:

Set up an example (combination of models + applicant cards), one participant – corresponding data card – analyse consequences of privacy to applicant then look into the others and discuss

Example discussion:

> Will Applicant name get the loan they have applied /which credit product will they have access to?
>
> How do you feel about this scenario?
>
> What do you feel is the impact of the implementation of differential privacy in this scenario?

Change model cards to create other examples

Part 3 - Post-game discussion

COMFORT BREAK

In general, what are your feelings regarding differential privacy?

Can you describe any positives / advantages that might come from DP being implemented?

Why do you think that?

Can you describe any negatives/ disadvantages that might come from DP being implemented?

Why do you think that?

Do you think it could benefit the applicants?

How would you feel if this technology was implemented?

In a scale of 1 to 5 where 1 is not confident and 5 is very confident where do you position yourselves in relation to your knowledge of the loan application process?

Why has your response changed? What has contributed to you moving from X to Y on the scale?

Probe questions... did the ppt have any impact? Has playing the game had any influence on your understanding? ...

How are your feeling about the current process of loans and what would you like it to be like?

D.3 Presentation

DIFFERENTIALLY PRIVATE CONSUMER CREDIT IMAGINARIES

Ana Rita Pena

PhD Candidate at the
Horizon Centre for
Doctoral Training

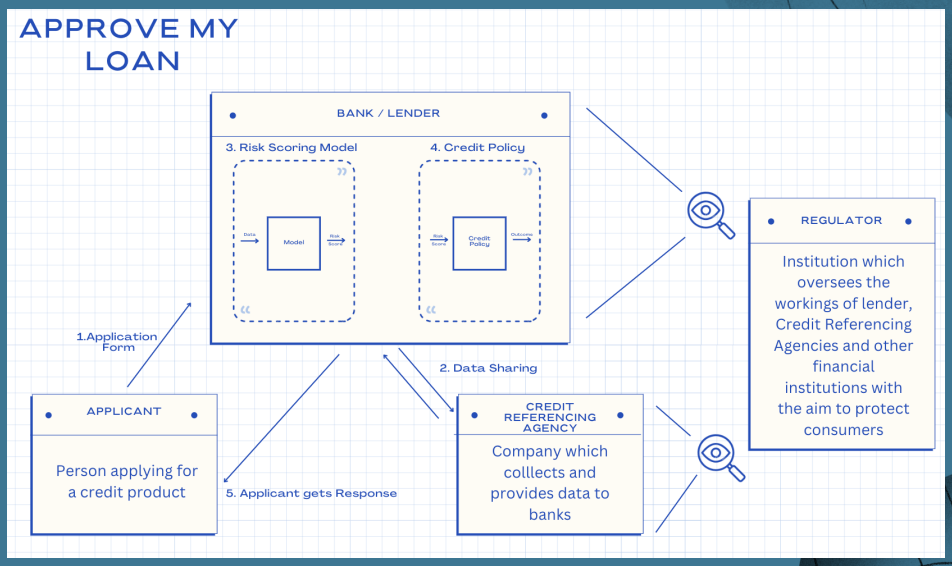
ana.pena@nottingham.ac.uk

Engineering and
Physical Sciences
Research Council

horizon
CENTRE FOR DOCTORAL TRAINING

University of
Nottingham
UK | CHINA | MALAYSIA

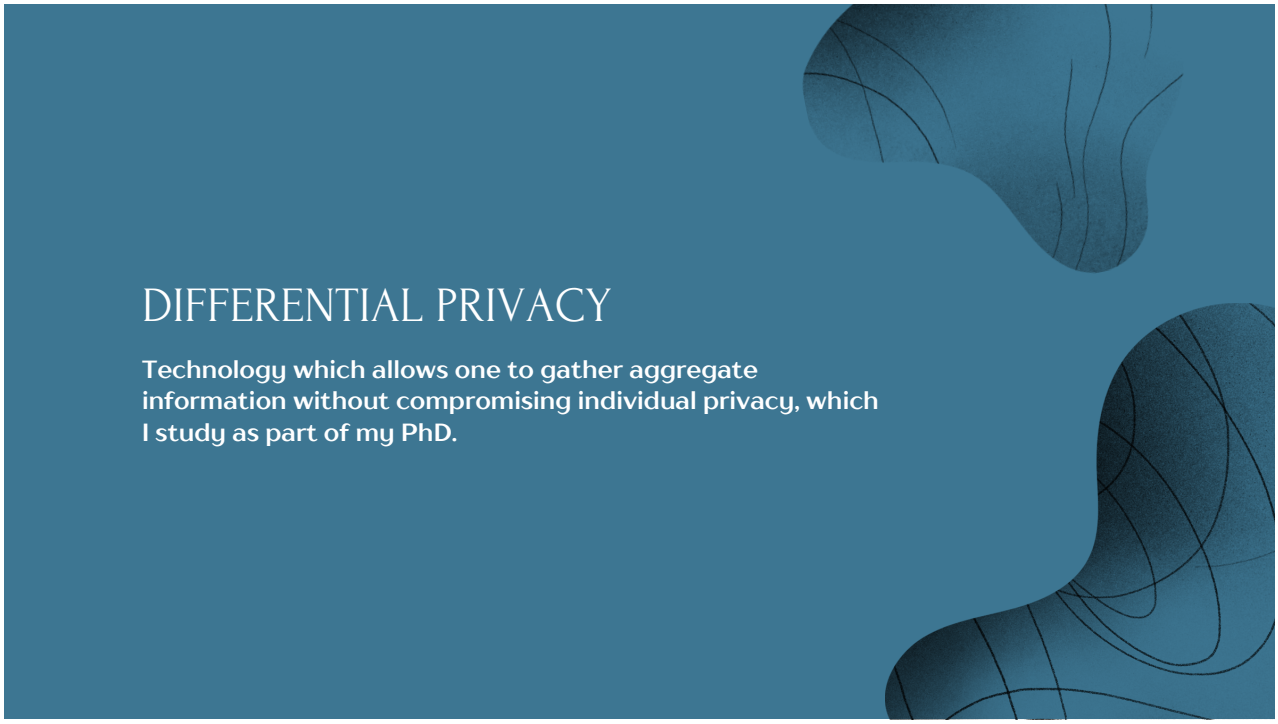
LOAN APPLICATION: BEHIND THE SCENES



QUESTIONS ?

DIFFERENTIAL PRIVACY

Technology which allows one to gather aggregate information without compromising individual privacy, which I study as part of my PhD.



WHY DIFFERENTIAL PRIVACY

Several controversies led to the creation of different techniques to enhance privacy.

However most privacy methods cannot prevent either reconstruction attacks (example in the image) or model inversion attacks.

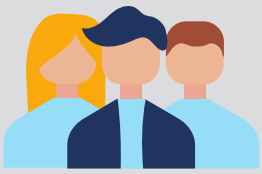
The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now

19 Pages • Posted: 4 Jun 2012 • Last revised: 3 Sep 2015

Daniel Barth-Jones
Columbia University - Mailman School of Public Health, Department of Epidemiology

Date Written: July 2012

CONCEPT INTRODUCTION

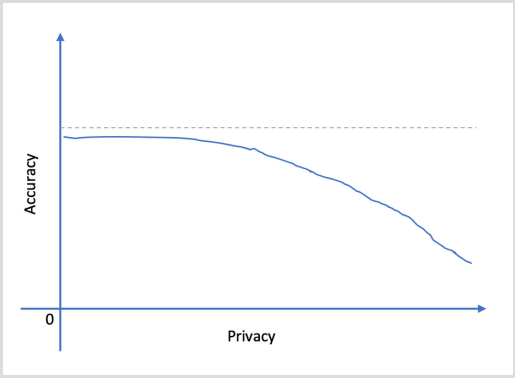


Differential Privacy works like looking through fogged glass.

The more fog there is on the glass the harder it is to identify and differentiate specific people out of a group.

PRIVACY - ACCURACY TRADE OFF

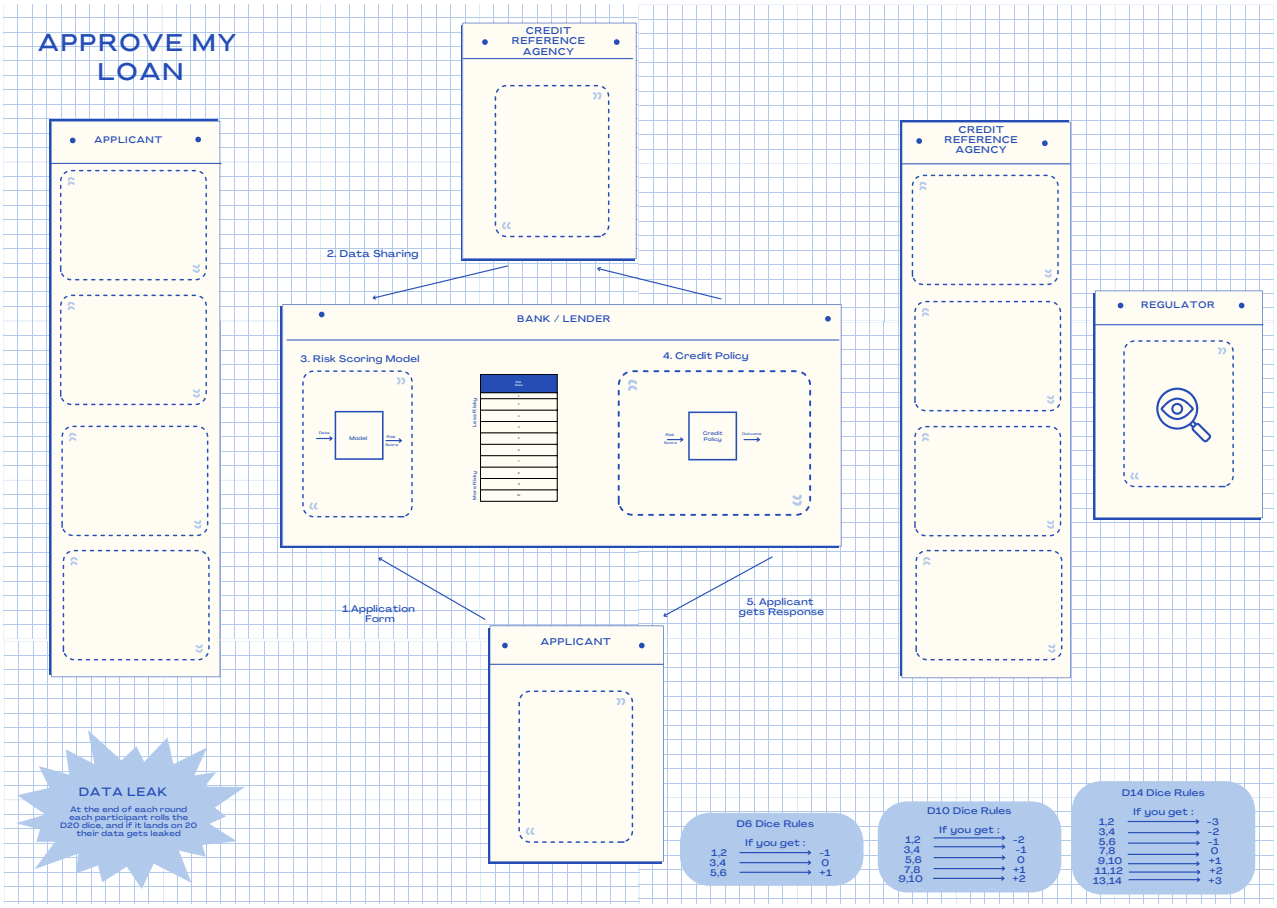
The privacy accuracy trade-off is defined by a privacy budget we can set.



QUESTIONS ?

D.4 Game Board

APPROVE MY LOAN



DATA LEAK

At the end of each round each participant rolls the D20 dice, and if it lands on 20 their data gets leaked

D6 Dice Rules

If you get :

1,2	-1
3,4	0
5,6	+1

D10 Dice Rules

If you get :

1,2	-2
3,4	-1
5,6	0
7,8	+1
9,10	+2

D14 Dice Rules

If you get :

1,2	-3
3,4	-2
5,6	-1
7,8	0
9,10	+1
11,12	+2
13,14	+3

D.5 Cards

DATA LEAK

An attacker was able to break through the security of a credit referencing agency and all data was leaked.
Through the bank account transaction data it was found that Rosie suffered from a chronic illness which affected her health insurance. It was found out that Mohamed gambled a few times a week and how much Adam earned and that one of Natasha's children is transgender.

APPLICANT

NATASHA

Natasha has recently gotten divorced, and she has two children, one of them transitioning using the private health system. Natasha has applied for a loan to cover medical expenses.

APPLICANT

ADAM

Adam is 40 years old and has been working in Data Science for the last 10 years. He has a wife and a daughter. He has a series of credit cards which he gets for the joining benefits and uses for the interest free period. He is looking to get another credit card.

APPLICANT

ROSIE

Rosie is a 70-year-old lady who has lived in Mansfield most of her life. She worked as a nurse for 30 years and retired five years ago. Rosie never married or had children. Rosie is looking into getting a credit card to help her spread out the cost of big purchases and help with the living cost.

APPLICANT

MOHAMED

Mohamed has just finished university and is waiting for his graduate program to start, before it starts Mohamed has applied for a small personal loan to pay for a holiday with his university friends.

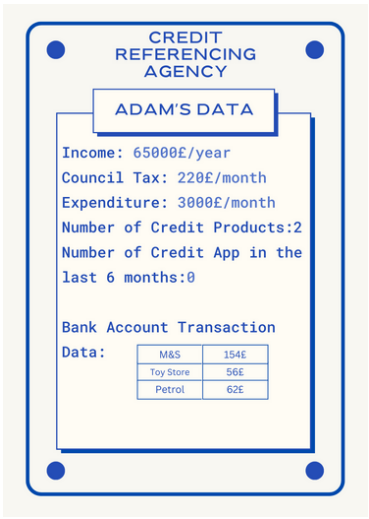
CREDIT REFERENCING AGENCY

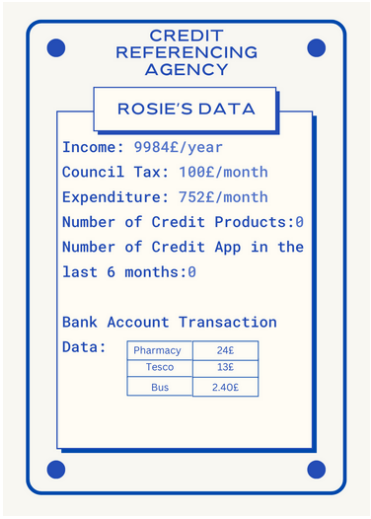
NATASHA'S DATA

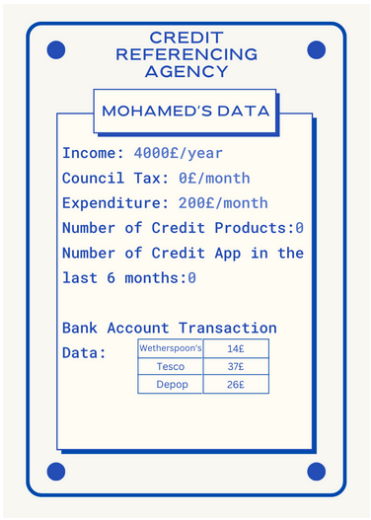
Income: 33000£/year
Council Tax: 220£/month
Expenditure: 1456£/month
Number of Credit Products:1
Number of Credit App in the last 6 months:0

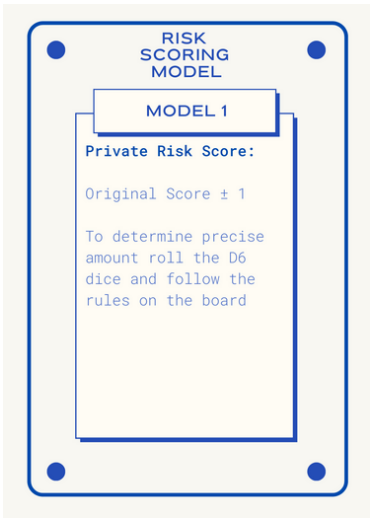
Bank Account Transaction Data:

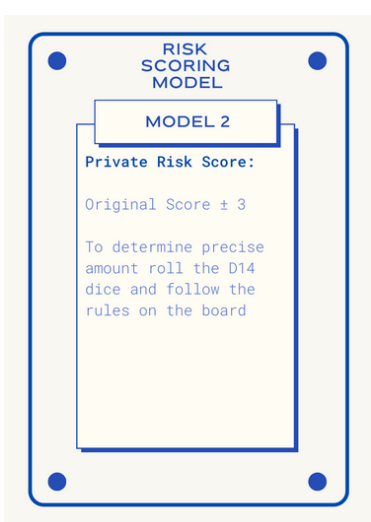
Pharmacy	36£
Lidl	121£
Petrol	62£

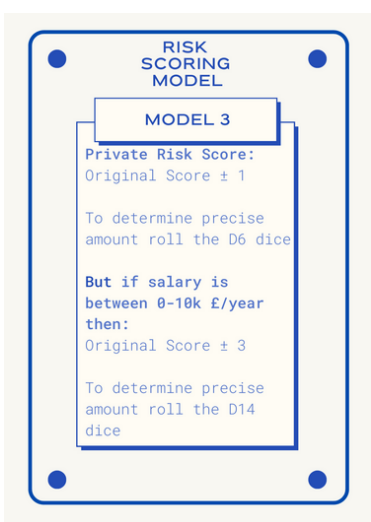

CREDIT REFERENCING AGENCY
ADAM'S DATA
Income: 65000£/year
Council Tax: 220£/month
Expenditure: 3000£/month
Number of Credit Products:2
Number of Credit App in the last 6 months:0
Bank Account Transaction Data:
M&S 154£
Toy Store 56£
Petrol 62£


CREDIT REFERENCING AGENCY
ROSIE'S DATA
Income: 9984£/year
Council Tax: 100£/month
Expenditure: 752£/month
Number of Credit Products:0
Number of Credit App in the last 6 months:0
Bank Account Transaction Data:
Pharmacy 24£
Tesco 13£
Bus 2.40£


CREDIT REFERENCING AGENCY
MOHAMED'S DATA
Income: 4000£/year
Council Tax: 0£/month
Expenditure: 200£/month
Number of Credit Products:0
Number of Credit App in the last 6 months:0
Bank Account Transaction Data:
Wetherspoon's 14£
Tesco 37£
Depop 26£


RISK SCORING MODEL
MODEL 1
Private Risk Score:
Original Score ± 1
To determine precise amount roll the D6 dice and follow the rules on the board


RISK SCORING MODEL
MODEL 2
Private Risk Score:
Original Score ± 3
To determine precise amount roll the D14 dice and follow the rules on the board


RISK SCORING MODEL
MODEL 3
Private Risk Score:
Original Score ± 1
To determine precise amount roll the D6 dice
But if salary is between 0-10k £/year then:
Original Score ± 3
To determine precise amount roll the D14 dice

RISK SCORING MODEL

MODEL 4

Private Risk Score:
Original Score ± 1

To determine precise amount roll the D6 dice

But if salary is between 50-100k £/year then:
Original Score ± 3

To determine precise amount roll the D14 dice

RISK SCORING MODEL

MODEL 5

Private Risk Score:
Original Score ± 2

To determine precise amount roll the D10 dice

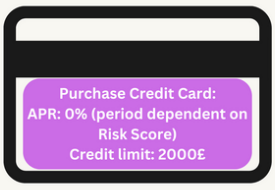
But if council tax is between 0-150 £/month then:
Original Score ± 1

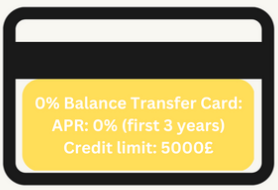
To determine precise amount roll the D6 dice

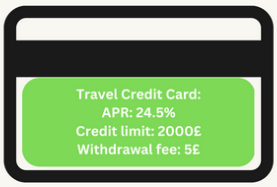
CREDIT POLICY

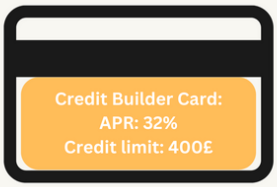
Original Risk Score:

Rosie	8
Mohamed	5
Adam	2
Natasha	6

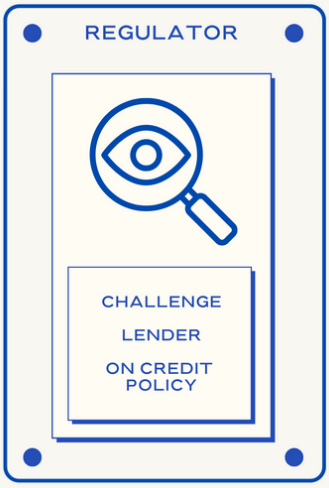





Travel Credit Card:
APR: 24.5%
Credit limit: 2000£
Withdrawal fee: 5£


Credit Builder Card:
APR: 32%
Credit limit: 400£


REGULATOR
CHALLENGE
CREDIT REFERENCING AGENCY
ON DATA COLLECTED


REGULATOR
CHALLENGE
LENDER
ON CREDIT POLICY


REGULATOR
CHALLENGE
LENDER
ON RISK MODEL