leaders

CASE STUDY

www.aileaders.eu

Biased AI models used for Credit Scoring

By
Rodrigo Cetina-Presuel
UPF Barcelona School of Management

Co-funded by the European Union

CONTENTS

Co-funded by the European Union

This project has been funded with support from the European Commission. The author is solely responsible for this publication (communication) and the Commission accepts no responsibility for any use may be made of the information contained therein. In compliance of the new GDPR framework, please note that the Partnership will only process your personal data in the sole interest and purpose of the project and without any prejudice to your rights.

www.aileaders.eu

| 01 Abstract

| Type of OER (Case study, Simulation, Scenario exercise, …)

This OER is a **case study on how biased or non-representative data can lead to undesirable outcomes, discrimination and inclusion through the use of automated credit scoring services.** The case study analyses the emergence and growth of automated credit decision systems that use algorithms to analyze various data points to assess an individual's creditworthiness to decide to approve or decline a credit application. The case focuses on how credit score systems work, using a simulation to illustrate what kind of data points they collect and what kinds of personal data they use in order to make decisions. The case study also raises awareness about the issues generated by biased or non-representative data sets and how they can make credit scoring systems unfair, functioning as barriers and not as improvements in the way citizens can access to credit and lending services in fair conditions.

| Goal or Purpose

The goal of this case study is to raise awareness of the potential of automated credit scoring systems to improve access to credit as well as of the challenges and the risks for privacy and discrimination the use of these systems entails, particularly when they use biased data sets and are trained with data that is not representative of the populations they are indented to serve and how this can exacerbate lack of access to credit in populations that already had trouble doing this to begin with.

| Expected Learning Outcomes

- The student will be able to **identify** ethical risks in credit scoring models and propose corrective measures.
- The student will **understand** the key ethical issues in AI applications for lending and credit scoring, including bias, discrimination, transparency, and data privacy.

| Suggested Methodological Approach

This case works best as problem-based learning in which instructors should guide a discussion with students once they have familiarized themselves with concepts related to access to credit, credit scoring services and contemporary credit application and lending practices. Topics for discussion, concerns and potential solutions are provided, but instructors should encourage students to think on their own and identify other potential concerns they may have. Examples and supplementary readings are provided via links as well as access to a simulation intended to illustrate what kind of personal data is used in automated credit assessment and what kinds of results these systems yield.

| Keywords

Lending, credit scoring, privacy and personal data protection, transparency, bias, discrimination.

02

INTRODUCTION TO THE CASE STUDY

Automated Access to Credit

Introduction: Automated Credit Scoring Systems

You may have seen several online offers from credit card companies that promise to approve a brand-new credit card (and corresponding line of credit) in as little as one or two days. Some promise to calculate your credit limit in a couple of minutes and authorize use "right away." Online automated credit applications have made the process easier and much quicker by calculating your credit score almost on the spot. This can provide companies with automated recommendations and/or approvals on a range of financial products, not only credit cards, but also other products such as personal loans or mortgages, while also calculating and making you an offer on the interest rates you must pay for each of these products.

Credit scores are computer-based models that correlate a series of factors with the probability that you may default on your debt payment. In other words, a credit score depicts your **credit worthiness** or your ability and reliability to repay any given loan. Credit scores are based on your credit history, built around a series of data points like the number of accounts you have and for how long they have been open, your total levels of debt, your repayment records, the types of loans you have and have received, the length of your credit transactions, the proportion of debt you are using and whether you applied for new accounts in a given period of time. By using these data, credit scores can separate good credit risks from bad ones and to classify would-be borrowers to predict their probability of default.

In the finance sector, banks and financial institution use credit scores to decide on loans or credit cards, insurance companies use it to evaluate the risk profile of policyholders, retailers use it for installment plans, car dealerships use them to assess eligibility, terms of auto financing and mortgage lenders use them to evaluate the creditworthiness of homebuyers in real state, and even your phone company uses them to decide if you are eligible for prepaid or postpaid plans or to determine if they can offer you a brand new phone to be financed and paid in installments.

Enter **automation**. The combination of multiple points of data and multivariate assumptions ae statistically combined and weighted to enable automated credit decisions. **Artificial intelligence** and **machine learning** employ statistical models and data analysis to make the credit assessment process faster and more streamlined. This allows for greater speed and potentially, greater accuracy and confidence in loan decisions. In the last few years, banking and credit industry have begun offering several **Fintech** (*financial technology)* products that offer automated credit decision-making solutions.

AI-driven credit decisions can potentially **improve efficiency and performance, reducing costs for financial institutions, which may also benefit consumers by expanding credit access or making credit less expensive:**[1] "The new high-performance models allow banks to define lending (and capital) parameters more precisely and thus sharpen their ability to approve creditworthy customers and reject proposals from customers who either are not creditworthy or cannot afford further debt. In fact, the banks (and fintech companies) that have put such new models in place have already increased revenue, reduced credit-loss rates, and made significant efficiency gains thanks to more precise and automated decisioning."[2]

However, there are several risks and drawbacks that we will explore during this case study. Among them, is the lack of **explainability** or inability to explain why programs make decisions. Another is **disparate impact** or discrimination that may affect certain protected groups more than others due to business practices (such as how one makes lending decisions). This, as we shall see, is directly related to **the use of flawed or biased data** to make credit decisions, which may end up **making access to credit more difficult** not easier to some collectives, thus **aggravating inequality in access to credit.**[3]

[1] https://www.congress.gov/crs-product/IF12399
[2] https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models
[3] https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit

03

DATA MODELS AND AI-DRIVEN CREDIT DECISIONS

| AI-Driven Credit

Specific Advantages

Lending decisions made with the aid of Artificial Intelligence have potential advantages that go beyond improvements in efficiency, and in driving down the costs associated with making credit available to people, making it less expensive. Potentially, Machine Learning in credit scoring can also expand the number and nature of the data points that factor in assessing people's credit worthiness and can offer the possibility of creating scores for communities that have been chronically underbanked and grant **access to banking and lending solutions for entire communities and individuals** that have traditionally been exclude from formal financial institutions.

By offering alternatives by incorporating a wider array of data, **AI-based scoring** "enables the evaluation of individuals without traditional credit histories by examining alternative data sources such as online transactions, social media interactions, browsing habits, or mobile app usage."[4] For those with traditional banking and financial records, AI-based scoring can provide more fine-tuned, enhanced analysis to make credit scoring and lending decisions more accurate. For those that lack traditional histories, it can offer the possibility of access to credit and other banking services.

Traditional Credit Score Data Sources	Alternative Credit Score Data Sources
• Credit cards and credit lines open • Automobile loans • Mortgages • Credit payment history • Credit inquiry histories • Bankruptcy filings	• Cash-flow data • Bill payments • Rental data • Employment records • Records on traffic violations or disputes • Telecom and mobile data usage patterns • Social media data • Behavioral data • Tax payments

Image 1. Examples of traditional and alternative credit data. Source: https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf

Currently, there is a large fraction of the world population that is considered what is known as **credit invisible** (no access) or **credit thin** (poor access in bad conditions) consumers. Just in the United States, it is estimated that 45 million consumers are unserved or underserved by traditional credit models.[5] In countries like India and South Africa, more than half of their populations have no good options for accessing credit.[6]

Where there was access, there is the potential to make it fairer and more efficient, where no access was possible, there is the potential to finally make it accessible, while also ensuring that this access is equally efficient and fair. Thus, there is great potential for **more inclusive and sustainable credit scoring models** which can in turn foster progress and financial security for a broader population, more people than ever before have the potential of having access to banking and financial services that may in turn give way to more competitive lending products in the market which may end up making credit more affordable. [7]

Why are we then worried that AI-Driven Credit Decisions could increase, not decrease, financial exclusion?

---

[4] https://www.orfonline.org/expert-speak/ai-and-credit-scoring-the-algorithmic-advantage-and-precaution
[5] https://newsroom.transunion.com/more-than-45-million-americans-are-either-credit-unserved-or-underserved---approximately-20-migrate-to-being-credit-active-every-two-years#_edn1
[6] https://svitla.com/blog/machine-learning-for-credit-scoring/
[7] https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf

Credit Scoring and the use of Personal Data

The Credit Approval Process

Credit scoring and credit approval involve complex processes that run behind the scenes in order to arrive a decision that involve inputs from the applicant, the bank or financial entity, a credit bureau and a credit scoring agency. You, and the other entities supply a wealth of data into the process that includes your personal data. Here is an image that roughly shows how the process works. As you can see, it not only maps the entities involved, it also shows you the layers of security used to protect the data used for the credit scoring and subsequent decision.

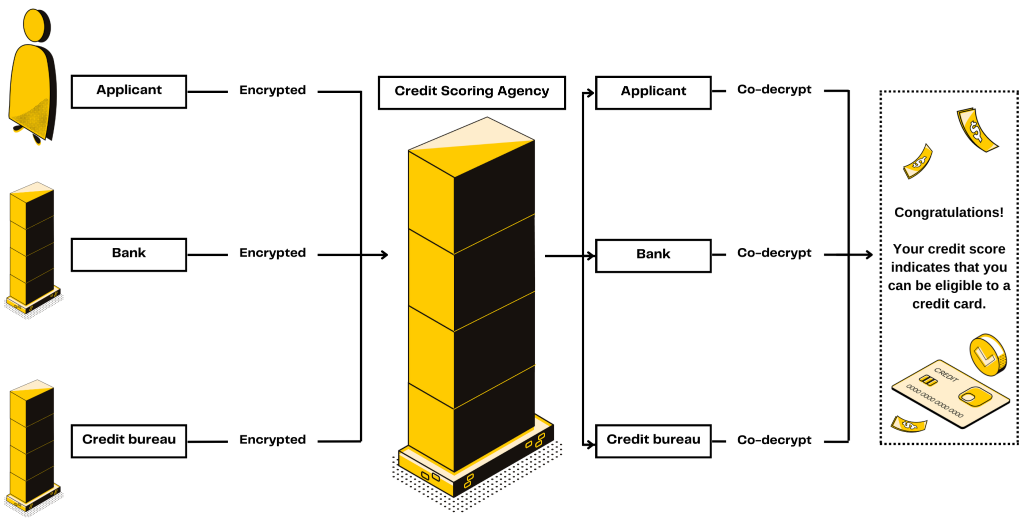


Image 2. A diagram that outlines the credit scoring and lending decision process. Source: https://huggingface.co/spaces/zama-fhe/encrypted_credit_scoring

To illustrate how this process works and what kinds of data you need to input in order for your credit worthiness to be assessed and a determination of giving you access to credit, such as a credit card, can be made, **click on the following image.** You will access a credit scoring simulation that will give you an assessment on your likelihood of getting a credit card approved.

04

RISKS AND CHALLENGES

| The Challenges and Risks of Biased and Non-Representative Data

Explainability Concerns

While Machine Learning models for credit scoring and lending offer several advantages as outlined before, there are also some serious concerns about transparency, fairness and the potential for bias reinforcement which in turn can lead to credit exclusion, not inclusion.

One specific concern is **lack of explainability** or the inability to explain my Machine Learning programs make particular decisions after analysing data inputs. This is a problem for the users of the systems, regulators and third parties who may not be able to understand and explain why a program did, and why. AI's ability to "react to large volumes of diverse inputs, beyond the reach of human cognitive ability – is also **ML's Achilles Heel** as **such complexity is often opaque in terms of the decision-making process that precedes a decision**."[8] Machine Learning that makes decisions that so complex that they are not easily interpreted or explained by humans[9] is usually referred to as **Black Box ML** or **Black Box AI.**

This is a significant challenge because it **limits transparency and accountability**. When an applicant is denied a loan, the lender should be able to explain the reason for the denial. If lenders cannot sufficiently explain how the automated systems arrived at a decision, then applicants may not have enough elements to contest the decision and may feel defenseless, undermining trust in the financial institution itself. Furthermore, lenders must make sure they can substantiate their decisions in order to comply with legislation that may mandate that credit and lending decisions must have clear motives to be disclosed to consumers.[10] Another problem is that **Black Box AI** gets in the way of efforts to improve system. If a decision is not adequate, "it is **extremely difficult to analyse why the mistake has been made**, or **to determine what needs to be done to correct the model**."[11]

Being able to explain AI decisions is also essential to gain the trust of users and to ensure that the decisions are fair and just which is why **methods to audit these systems become essential.** Explainability should usually focus on explaining why "this particular input lead(s) to that particular output"[12] but something that is also essential is knowing what internal data forms the structures of a particular program. In the following section, we focus on issues related to the data used to train systems which is what is ultimately combined with the particular data submitted by a user and enables the AI system to generate a particular decision.

Lack of Diversity of Datasets and Flawed Data

As explained, biased data and biased algorithms can make automated decision-making lead to outcomes that put those that have trouble accessing good credit, or even financial services to begin with, at a disadvantage. Minorities and low-income groups usually suffer from these biases in a disproportionate manner.

However, research shows that this is not the only problem. Different outcomes for minorities and majorities are not just related to bias, but to the fact that minority and low-income groups have less data in their credit histories because they are usually underrepresented in access to credit to begin with.[13]: "This means that when this data is used to calculate a credit score and this credit score (is) used to make a prediction on loan default, then that prediction will

[8]https://www.kcl.ac.uk/challenges-of-ai-explainability
[9]https://www.sciencedirect.com/science/article/pii/S0828282X21007030
[10]https://www.congress.gov/crs-product/IF12399
[11]https://www.kcl.ac.uk/challenges-of-ai-explainability
[12]https://arxiv.org/pdf/1806.00069
[13]https://arxiv.org/abs/2105.07554

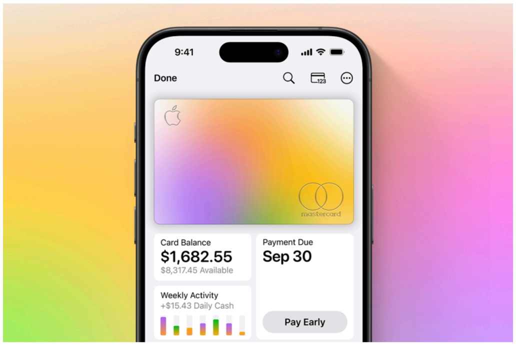
be less precise. It is this lack of precision that leads to inequality, not just bias.[14]

The lack of diversity in data sets to train machine learning models results in harms for specific communities, increasing inequality. Systemic inequalities persist in dataset curation and inequality of access. In some fields, it is also because there is unequal opportunity to participate in the building of those datasets.[15] The performance of any AI system is heavily determined by the datasets it analyses using statistics because their outcomes come from identifying patterns in the data: "the quality of an AI system's underlying dataset is crucial for its effectiveness." [16]vWithin the context of lending and access to credit: "**It's a self-perpetuating cycle… We give the wrong people loans, and a chunk of the population never gets the chance to build up the data needed to give them a loan in the future.**"[17]

The Risk of Algorithmic Bias

While one of the advantages of automated analysis of credit applications may be a reduction in subjectivity in the decision-making process for granting a loan, there is a risk that these processes consolidate "existing bias and prejudice against groups defined by race, sex, sexual orientation, and other attributes"[18] some of which are special categories of personal data protected by the law.[19] This is because datasets usually contain past decisions made from financial institutions or because they do not contain sufficient data about certain groups which could even lead to their discrimination.

If you click on the following image, you will be able to access a news article that explains how **Apple Card's** automated credit card approval system, provided by **Goldman Sachs,** led to complaints of alleged discrimination against female applicants that claimed they were offered lower credit limits or denied a card, even if their husbands got approvals and better rates.



Apply to see your credit limit offer.

Denial of access to lending markets can be discriminatory when decisions are skewed based on those pre-existing biases. It can also manifest itself because of differential treatment that offers poorer conditions for those that suffer from discrimination, such as different, less advantageous fees, or higher interest rates.[20]

[14]https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/
[15]https://pmc.ncbi.nlm.nih.gov/articles/PMC10667100/
[16]https://www.human-technology-foundation.org/news/diversity-in-ai-towards-a-problem-statement
[17]https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/
[18]https://link.springer.com/article/10.1007/s00146-023-01676-3
[19]https://gdpr-info.eu/art-9-gdpr/
[20]https://link.springer.com/article/10.1007/s00146-023-01676-3

Bias in datasets may **erect further barriers for access to financial services for traditionally underserved populations** that had trouble accessing these services to begin with. Traditional credit scoring systems penalize those without formal credit histories, but even systems that use non-traditional data, if they exhibit biases in their decisions or reinforce the biases of their creators, may **reinforce exclusion** disparately impacting those underserved communities.

Limited access to traditional financial services by certain collectives may perpetuate wealth disparities, past discrimination in this field has led to long-term effects in the overall generational wealth of certain groups, with consequences such as "limited home ownership, reduced entrepreneurial opportunities, and generational wealth gaps."[21]

If not carefully designed and implemented, **AI models will "perpetuate inequities"** [22]negating any possible positive impact from a decision-making process that reduces subjectivity only on paper, as it carries with it the structural biases already present in society.

Where do biases come from?

Algorithmic biases may come from different sources, **including biased training data** which may be historical data that reflects existing societal biases. This is known as "bias in, bias out" or when a model is trained on data related to already biased outcomes. AI is just likely to replicate the issues of the past.

Lack of diversity issues are not only present in the data. They may also appear in the **product design teams as well:** "homogeneity among data scientists and developers contributes to the perpetuation of bias in AI systems."[23] Lack of experiences or understanding of underserved communities may make those that design the systems and supply the training data blind to their struggles and challenges or specific situations, which may lead to inaccurate analysis of their credit histories.

Biased decisions **may only reinforce and entrench themselves over time** as the systems make more and more incorrect assessments that are then deemed as valid, generating **feedback loops** where incorrect denials of credit for a specific collective are reflected in future training data, **further perpetuating inequality and unfairness.** [24]

[21]https://ijsra.net/sites/default/files/IJSRA-2024-2257.pdf
[22]https://ijsra.net/sites/default/files/IJSRA-2024-2257.pdf
[23]https://ijsra.net/sites/default/files/IJSRA-2024-2257.pdf
[24]https://ijsra.net/sites/default/files/IJSRA-2024-2257.pdf

05

GUIDELINES FOR INSTRUCTORS

| GUIDELINES FOR INSTRUCTORS

On the Case Study

This case study provides instructors with the possibility to enable a class discussion about several topics related to **financial inclusion and exclusion**, among them the **potentials and risks of automated credit scoring and automated credit decisions.**

First, you could start with a **discussion on AI-driven credit scoring and AI-driven financial services**, discussing the **merits of automation** and the **potential** of using different types of **data sets to service different populations and focus on inclusion.**

After learning about the **credit approval process**, you should **focus on the use of personal data for these kinds of decisions**, since AI-driven models require not only datasets for training, but also for users to input personal data as demonstrated by **the simulation in page 8 of this case study.**

Here are some prompts for your students:

- Do you think there are any **risks to people's privacy** and **personal data protection**?
- If yes, why do you think it **is important to protect personal data**?
- What **consequences** could clients face if their **data is out in the open**?
- Besides data leaks and other cybersecurity threats, **what other risks can you associate with the use of personal data to make lending decisions**?

Once you enter part four you **should discuss the challenges with your students**, one by one and **then focus on solutions**. The case study includes some references and further readings in the annexes below, but an interesting exercise is to assign students in groups and have them do some research to **come up with possible solutions for each of the issues presented.** One or more groups can focus on research on explainability concerns, others on solutions related to the problem of diversity in data sets and others on algorithmic biases in general. After doing some research, you can ask them to **make a brief presentation to present the solutions they could implement.**

You can instruct them to **search for solutions that range from regulatory and legislative methods to more technical solutions.** There are great resources that speak about these topics online, so it is also **a good opportunity to have them conduct some research where they think critically** and **identify reliable sources** that propose **worthwhile solutions.**

Finally, if you think it would fit within the context of your class, you can **discuss the benefits and pitfalls of the general trend of automating decision-making that can impact people's lives and the need to at least have humans supervising these decisions.** You can **articulate an interesting discussion around the need of human interaction and granting opportunities to credit applicants vs. the merits of decisions aided by automation** or even delegated to AI systems and **the potential of decisions that eliminate subjectivity**. Is this desirable? Is this even possible? Or, on the contrary, is some degree of subjectivity and human agency necessary for fair decisions?

06

ANNEXES

Further reading

- Alternative Data for Credit Scoring: https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf

- Should AI decide Who gets a Loan? https://odsc.medium.com/should-ai-decide-who-gets-a-loan-83c6f259081b#:~:text=Types%20of%20Data%20Used%20in,data%20point%20analyzed%20is%20collateral.

- ML applied to Credit Risk: Building explainable models: https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2022/10/Idanae-3Q22.pdf

- Accuracy of Explanations of Machine Learning Models for Credit Decisions: https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/22/Files/dt2222e.pdf

- From Inherent Racial Bias to Incorrect Data – The Problems with Current Credit Scoring Models: https://www.bbc.com/news/business-50365609

- Bias isn't the only Problem with Credit Scores – And no, AI can't help: https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/

- How Flawed Data Aggravates Inequality in Credit: https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit

- Apple's 'Sexist' Credit Card Investigated by US Regulator: https://www.bbc.com/news/business-50365609

- Designing Next-Generation Credit-Decisioning Models: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models