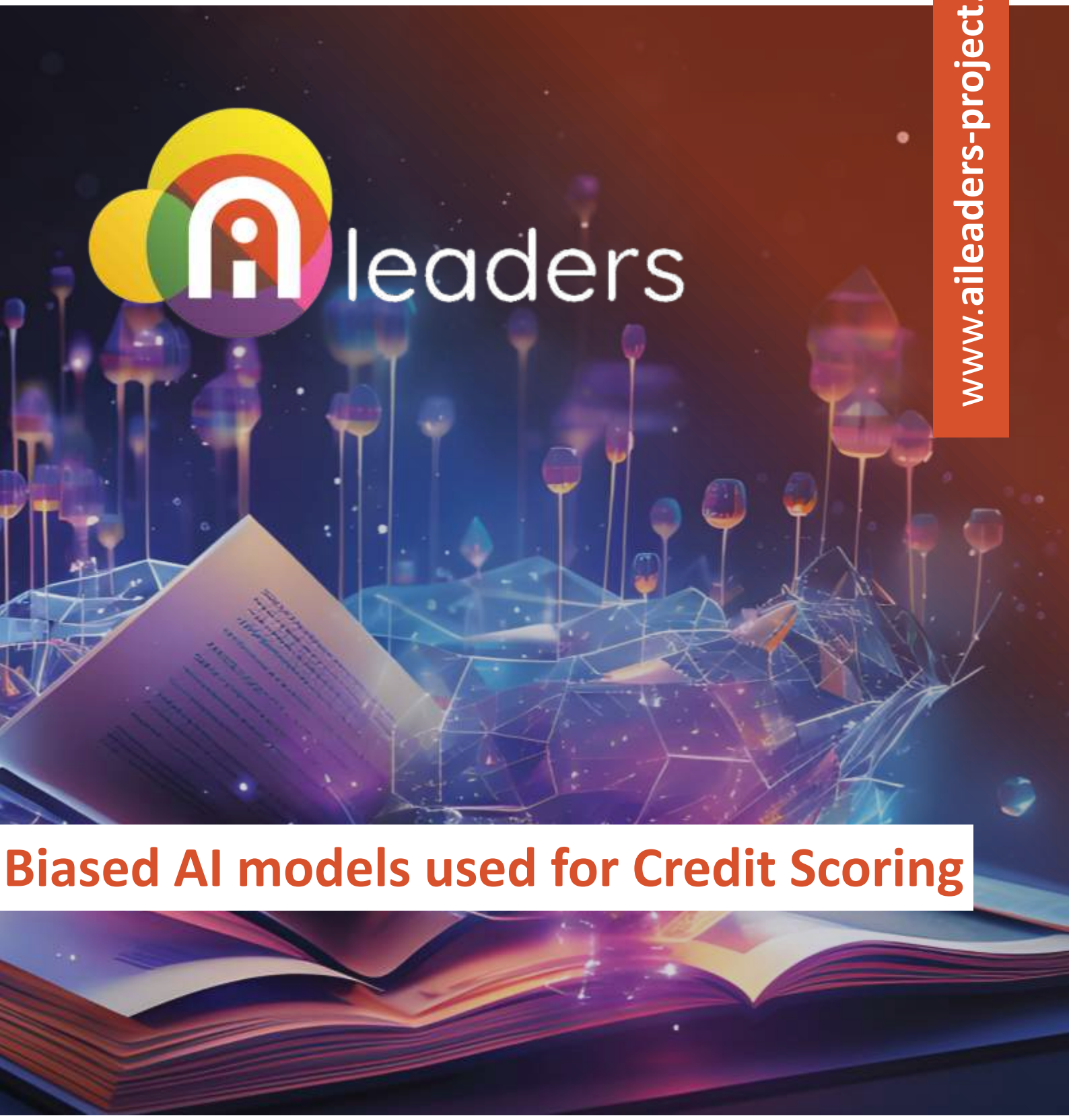
www.aileaders-project.eu



Biased AI models used for Credit Scoring

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

CASE STUDY: 01

This OER is a case study on how biased or non-representative data can lead to undesirable outcomes, discrimination, and exclusion 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 analyse 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 datasets and how they can make credit scoring systems unfair, functioning as barriers and not as improvements in the way citizens can access credit and lending services in fair conditions.

Goal/Purpose

The goal of this case study is to raise the awareness of the potential of automated credit scoring systems to improve access to credit as well as of the challenges and risks for privacy and discrimination that the use of these systems entails, particularly when biased datasets are used and also when the models are trained with data that is not representative of some populations.

Expected Learning Outcomes

01 The student will be able to **identify** ethical risks in credit scoring models and propose corrective measures.

02 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

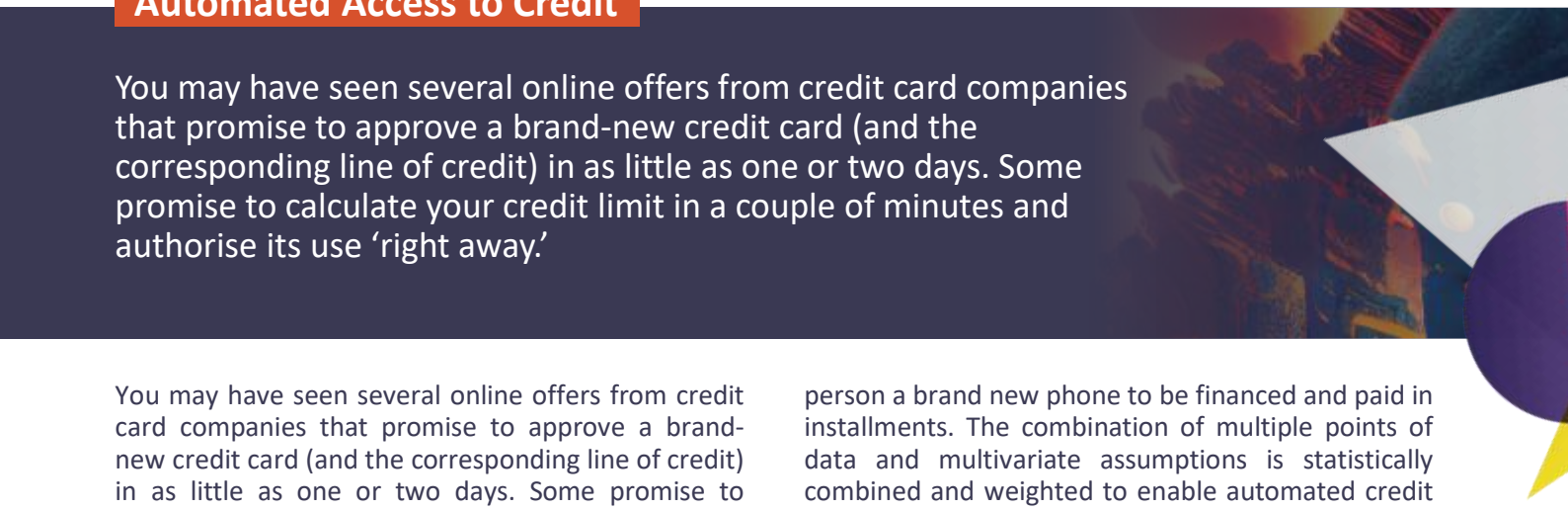
This case works best as problem-based learning in which instructors should guide a discussion with students once they have familiarised themselves with concepts related to access to credit, credit scoring services, as well as contemporary credit application and lending practices. Topics for discussion, concerns, and potential solutions are provided, but the instructors should encourage the 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

Automated Access to Credit

You may have seen several online offers from credit card companies that promise to approve a brand-new credit card (and the 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 authorise its use 'right away.'



You may have seen several online offers from credit card companies that promise to approve a brand-new credit card (and the 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 authorise its use 'right away.' Online automated credit applications have made the process easier and much quicker by calculating one's 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 an offer on the interest rates one must pay for each of these products.

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

In the finance sector, banks and financial institutions use credit scores to decide on loans or credit cards; insurance companies use them to evaluate the risk profile of policyholders; the retailers use it for installment plans; car dealerships use them to assess eligibility and terms of car financing; and mortgage lenders use them to evaluate the creditworthiness of homebuyers in real state. Even the phone company uses them to decide if one is eligible for prepaid or postpaid plans or to determine if they can offer a person a brand new phone to be financed and paid in installments. The combination of multiple points of data and multivariate assumptions is 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 both 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 programmes make decisions. Another one is disparate impact or discrimination that may affect certain 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 result in making access to credit more difficult and not easier to some collectives, thus aggravating inequality in access to credit [3].

1 Congressional Research Service. (2023). *Generative artificial intelligence: Overview, issues, and questions for Congress* (CRS Report No. IF12399). https://crsreports.congress.gov/product/pdf/IF/IF12399

2 Dash, R., Kremer, A., & Petrov, A. (2021). *Designing next-generation credit-decisioning models*. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models

3 Andrews, E. L. (2021, August 6). *How flawed data aggravates inequality in credit*. Stanford HAI. https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit

Data Models & 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 as well as 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 excluded 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 • behavioural 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 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 such as India and South Africa, more than half of the population have no good options for accessing credit[6]. Where there was access, there is the potential to make it fairer and more efficient, while 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.

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].

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

4 Bag, S. (2024). AI and credit scoring: The algorithmic advantage and precaution. Observer Research Foundation. https://www.orfonline.org/expert-speak/ai-and-credit-scoring-the-algorithmic-advantage-and-precaution

5 *TransUnion. (2022). More than 45 million Americans are either credit unserved or underserved; Approximately 20% migrate to being credit active every two years. TransUnion*. 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 *Svitla Team. (2024). Machine learning for credit scoring: Benefits, models, and implementation challenges. Svitla Systems.* https://svitla.com/blog/machine-learning-for-credit-scoring/

7 *Alliance for Financial Inclusion. (2025). Alternative data for credit scoring*. 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 for a decision to be made

This includes inputs from the applicant, the bank or financial entity, a credit bureau, and a credit scoring agency. A person and the other entities supply a wealth of data into the process that includes one's personal data. Here is an image that roughly shows how the process works. As can be seen, it not only maps the entities involved, but also shows the layers of security used to protect the data employed 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 one needs to input in order for their credit worthiness to be assessed and in order for the determination of providing access to credit, such as a credit card, to be made**, click on the following image**. A credit scoring simulation will be demonstrated that will provide an assessment of one's likelihood of getting a credit card approval.

The Challenges and Risks of Biased and Non-Representative Data

Explainability Concerns

While Machine Learning models for credit scoring and lending offer several advantages, 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 **the lack of explainability** or the inability to explain why Machine Learning programmes 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 programme 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 so complex that they are not easily interpreted or explained by humans is usually referred to as **Black Box ML** or **Black Box AI.**[9]

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 the 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 the 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 for gaining the trust of users and for ensuring 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 programme. In the following section, we focus on issues related to 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.

8 *King's College London. (n.d.). The challenges of AI explainability.* https://www.kcl.ac.uk/challenges-of-ai-explainability
9 *Kosinski, M. (2024, October 29). What is black box AI and how does it work? IBM*. https://www.ibm.com/think/topics/black-box-ai
10 *Congressional Research Service. (2023). Generative artificial intelligence: Overview, issues, and questions for Congress (CRS Report No. IF12399).* https://crsreports.congress.gov/product/pdf/IF/IF12399
11 *King's College London. (n.d.). The challenges of AI explainability*. https://www.kcl.ac.uk/challenges-of-ai-explainability
12 *Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. ArXiv*. https://arxiv.org/abs/1806.00069

The Lack of Diversity of Datasets and Flawed Data

As explained earlier, 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 related merely to bias, but also 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 be less precise. It is this lack of precision that leads to inequality, not just bias.[14] The lack of diversity in datasets to train machine learning models results in harms to specific communities, which increases inequality. Systemic inequalities persist in dataset curation and the 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] Within 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 the 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 by financial institutions, or because they do not contain sufficient data about certain groups which could 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.

13 *Blattner, L., & Nelson, S. (2021). How costly is noise? Data and disparities in consumer credit. ArXiv. https://arxiv.org/abs/2105.07554*

14 *Heaven, W. D. (2021). Bias isn't the only problem with credit scores – and no, AI can't help. MIT Technology Review. https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/*

15 *Arora, A., Alderman, J. E., Palmer, et al. (2023). The value of standards for health datasets in artificial intelligence-based applications. BMJ Health & Care Informatics, 30(1), e100888. https://doi.org/10.1136/bmjhci-2023-100888*

16 *Boch, A., & Kriebitz, A. (2023, September 29). Diversity in AI - Towards a problem statement. Human Technology Foundation. https://www.human-technology-foundation.org/news/diversity-in-ai-towards-a-problem-statement*

17 *Heaven, W. D. (2021). Bias isn't the only problem with credit scores – and no, AI can't help. MIT Technology Review. https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/*

18 Garcia, A. C. B., Garcia, M. G. P., & Rigobon, R. (2024). Algorithmic discrimination in the credit domain: What do we know about it?. AI & Society, 39, 2059–2098. https://doi.org/10.1007/s00146-023-01676-3

19 Article 9 of the General Data Protection Regulation: https://gdpr-info.eu/art-9-gdpr/

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 to those that suffer from discrimination, such as different, less advantageous fees, or higher interest rates.[20]

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 penalise 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.

The lack of diversity issues are present not only in the data. They may also appear in **product design teams:** "Homogeneity among data scientists and developers contributes to the perpetuation of bias in AI systems." [23] The lack of experiences with 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]

20 *Garcia,, A. C. B., Garcia, M. G. P., & Rigobon, R. (2024). Algorithmic discrimination in the credit domain: What do we know about it?. AI & Society, 39, 2059–2098. https://doi.org/10.1007/s00146-023-01676-3*

21 *Nuka, T. F., & Osedahunsi, B. O. (2024). From bias to balance: Integrating DEI in AI-driven financial systems to promote credit equity. International Journal of Science and Research Archive, 13(2), 1189–1206. https://doi.org/10.30574/ijsra.2024.13.2.2257*

22 *Nuka, T. F., & Osedahunsi, B. O. (2024). From bias to balance: Integrating DEI in AI-driven financial systems to promote credit equity. International Journal of Science and Research Archive, 13(2), 1189–1206. https://doi.org/10.30574/ijsra.2024.13.2.2257*

23 *Nuka, T. F., & Osedahunsi, B. O. (2024). From bias to balance: Integrating DEI in AI-driven financial systems to promote credit equity. International Journal of Science and Research Archive, 13(2), 1189–1206. https://doi.org/10.30574/ijsra.2024.13.2.2257*

24 *Nuka, T. F., & Osedahunsi, B. O. (2024). From bias to balance: Integrating DEI in AI-driven financial systems to promote credit equity. International Journal of Science and Research Archive, 13(2), 1189–1206. https://doi.org/10.30574/ijsra.2024.13.2.2257*

Guidelines for Instructors

On the Case Study

This case study provides instructors with the possibility to enable 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 **datasets to service various 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 datasets not only for training, but also for users to input personal data as demonstrated by **the simulation on 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 one 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 datasets, and still others on algorithmic biases in general. After doing some research, you can ask the students to **make a brief presentation to present the solutions they could implement.**

You can instruct the students to **search for solutions that range from regulatory and legislative methods to more technical solutions.** There are great online resources that speak about these topics, 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 as well as 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?

Further Reading

- Alliance for Financial Inclusion. (2025). Alternative data for credit scoring. https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf
- Alonso, A., & Carbó, J. M. (2022). Accuracy of explanations of machine learning models for credit decisions (Working Paper No. 2222). Banco de España. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/22/Files/dt2222e.pdf
- Blattner, L., & Nelson, S. (2022). How flawed data aggravates inequality in credit. Stanford Graduate School of Business. https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit
- Dash, R., Kremer, A., & Petrov, A. (2021). Designing next-generation credit-decisioning models. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models
- Heaven, W. D. (2021). Bias isn't the only problem with credit scores – and no, AI can't help. MIT Technology Review. https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/
- Kelion, L. (2019). Apple's 'sexist' credit card investigated by US regulator. BBC News. https://www.bbc.com/news/business-50365609
- Mensalvas, E., Guzmán, M. A. & Ruiz Bonilla, S. (2022). ML applied to credit risk: Building explainable models. https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2022/10/Idanae-3Q22.pdf
- ODSC. (2022). Should AI decide who gets a loan? Medium. https://odsc.medium.com/should-ai-decide-who-gets-a-loan-83c6f259081b