The Emerging Role of Social Scoring Systems in Today Society: State of the Art and Legal Issues

Carraro Marco[1], Debole Nicola[1], Stefan Massimo[1]
[1]Artificial Intelligence Systems, Università degli studi di Trento

Abstract

This project focuses on the introduction of Social Scoring systems within our society and their implications from a legal and moral perspective. The common thread running through the entire project is that the description of such systems necessarily leads to a continuous balance and contrast between the ideal benefits which they can bring, and the potential downsides they may conceal. In our research we verified that, in our opinion, some current uses of Social Scoring systems do not turn out to meet the positive expectations for which they were created. The project also highlights the state of the art of today's systems within various governments around the world and in some well-known private companies. We focused on a few actual case studies in order to make our research more concrete to then explain how the European Union addresses this issues. For this reason, we also found it necessary to list fundamental rights, cited within the major constitutions and declarations of rights, that were in possible opposition to the operation of Social Scoring systems. We then tried to tackle the open dilemma of the existence of a system built to "classifying" citizen behavior, centered on the use of Artificial Intelligence, that does not create a discriminatory system. Without presuming to be able to give a definitive answer to this question, we have tried to propose a possible solution to the problem.

Keywords: artificial intelligence; social scoring; social credit; AI act; rights; legal issues;

Contents

1 Introduction

1.1 Social Scoring

In this paper we will refer to Social Scoring as an umbrella term intended to define all those practices used to assign a particular score to a section of the population, through tracking systems and Artificial Intelligence. Within our research it will be emphasized how systems of this type are already being used, or partially used, in various places around the world. This, of course, does not allow to define in detail the structure of a Social Scoring system, since their uses and effects may differ from each other.

In general, it is possible to establish some fixed and recurring points of Social Scoring systems, including:

1. a **system of control**, which in our research we will refer to as authority, represented mainly by a government (local or state) but also by private entities, in the latter case mainly for commercial purposes;
2. a predefined part of the **population**, subject to tracking and recognition controls by an authority or private entity
3. a **ranking and evaluation system**, responsible for assigning scores to subjects involved in it. The evaluation methodologies are singular to each system and based on several possible phenomena. These methodologies are implemented through algorithms that based on the trend and prediction of a future trend related to a particular subject. The scores assigned can be either positive, and thus increase the score of a subject involved, or negative, going instead to decrease the subject's score;
4. a predetermined set of **benefits and advantages** for the most virtuous individuals within the system, contrasted with another set of disadvantages and penalties for those in the lowest ranking areas;

1.2 Applied contexts

In Social Scoring systems data is collected on personal behaviors and circumstances, such as expressions and contacts in social networks, consumer behavior, leisure activities, and movement profiles. What is special about social scoring is not only the breadth of the data used but also the breadth of the score's possible areas of application. The findings from this provide a detailed picture of the individual and allow conclusions to be drawn not only about solvency but also about reliability in all aspects of life. In this respect, the reliability criterion stands alongside the classic assessment of creditworthiness (Maamar, 2018).

Below we list the main areas of application of social scoring systems currently in place or under development:

1. Credit and financial reliability of a customer (as in Chinese SCS) (Liang & Chen, 2022);
2. Determine subjects which should be targeted for access, eligibility, suspicion or exclusion;
3. Transformation of people into automation and statistical analysis (CHENEY-LIPPOLD, 2017);
4. Surveillance practices (e.g. border control using emotional AI) (Sánchez-Monedero & Dencik, 2022);

As mentioned in the previous paragraph, the possibilities of applying a Social Scoring system can be the most varied. In particular, their function is highly dependent on the authority in control of the system itself. Surveillance and security systems, such as those related to the area of financial reliability are mainly used at the government level, whether local, national or international. The use of Social Scoring to transform physical subjects into statistical predictions, instead, may also predominantly involve private companies.

As will be explained in more detail within our research, the geographical location in which these systems are applied is a key aspect. This is because national and international organizations do not follow the same guidelines on the Artificial Intelligence front, thus causing a sharp contrast between products that could potentially already be in the market but turn out not to be compliant with the standards or regulations in place in that particular country.

1.3 Social Scoring goal

Underlying the question of what is the goal of social scoring, in our view it is fair to ask who wishes to benefit from the development of these systems. Considering, in fact, a utopian reality in which each subject is responsible and aware of his or her every action, social scoring systems would simply find no utility, since they would have no reason to control an environment that already functions perfectly, within which each agent "acts" in the best way. Having deduced that the above reality is not about the world in which we live, here is where social scoring systems can provide a supporting tool. At this moment the question arises: does such support benefit those who are an integral part of the system (the citizens) or those who control the system itself (the authority)?

Splitting and sorting citizens into categories potentially shape social life and redistribute resources, thereby disciplining individual behavior towards certain desirable goals defined by the authority. For example, the political logic of the Chinese SCSs is that local governments attempt to encourage social morality and enforce social management. This goal is further supported by the reward and punishment mechanisms imposed on the SCSs. To receive benefits and avoid punishments, citizens are motivated to participate in activities that can raise their

credit scores. In fact, the SCSs explicitly encourage people to get extra points by engaging in charitable donations, volunteer services, blood donations, and other social charitable activities, which are often not necessarily trustworthy or creditworthy related behaviors. Citizens are also expected to exercise self-discipline for maximizing their credit scores and be in conformity with the norms. Therefore, the point accumulation could stimulate people's engagement with social charities and law-abiding actions.

While the system can encourage self-discipline and self-restraint, they may not be able to enforce internalizing behaviors. This raises questions about the intrinsic quality of citizens, as people are mainly motivated by external regulations and incentives to enact trust-keeping activities (Liang & Chen, 2022).

Considering this, it is important to ask whether the utopian goal of a Social Scoring system is destined to remain so or, in the event that they become an integral part of everyday life, they can inherently change the habits of those involved, thereby creating a society that aims to improve itself more and more.

1.4 Social Scoring Systems: tackled EU values

The use of Social Scoring systems, of course, brings with it the need to think about its morality and the implementation of certain guidelines that can control and limit its use, to protect those directly affected, the citizens.

In this respect, the European Union has not been indifferent to the operation of these technologies, and has therefore classified the most of this kind of systems as "high-risk" systems. This means that under no circumstances can the results produced by algorithms operating in high-risk systems be considered as totally reliable, without having received the necessary supervision of a physical person.

Social scoring systems, such as those launched in China that track the trustworthiness of people and businesses, are classified as "contravening the Union values" and are going to be banned. The proposal also wants to prohibit AI systems that cause harm to people by manipulating their behavior, opinions or decisions; exploit or target people's vulnerabilities; and for mass surveillance.

But the rules carve out an exception allowing authorities to use the tech if they're fighting serious crime. The use of facial recognition technology in public places, for example, could be allowed if its use is limited in time and geography. The Commission said it would allow for exceptional cases in which law enforcement officers could use facial recognition technology from CCTV cameras to find terrorists, for example. The exception is likely designed to appease countries like France, which is keen to integrate AI into its security apparatus, but is opposed by privacy hawks and digital rights activists who have lobbied hard for these uses to be banned outright.

The EU is also keen to avoid issues of racial and gender bias, which have plagued the development of the technology from its inception. One of the Commission's requirements in the draft are that data sets do not "incorporate any intentional or unintentional biases" which may lead to discrimination.

The draft also proposes creating a European Artificial Intelligence Board, comprising one representative per EU country, the EU's data protection authority, and a European Commission representative. The board will supervise the law's application and share best practices (Heikkila, 2021).

2 State of the art

When thinking about Social Scoring, China is the first country that instantly comes to mind, mainly because of the media extensively covering its controversial political decision to use Social Scoring on its citizens. In fact, everyone knows that China is the country that most centered its monitoring through AI, deep learning, biometric identifications, etc. However, it is not the only entity that is imposing social monitoring solutions. For example, virtually every modern state applied in a way or another a system to mitigate the spread of COVID-19, providing a decent case for considering social scoring systems. In the following subsections different applications will be listed not only for countries, but even private companies, in order to understand the differences in methods and goals, dwelling on the pros and cons.

2.1 Countries

In this section we are going to describe different applications of the Social Scoring and Credit system, employing peculiar cases like UK, Italy and China. While in China they are already being implemented, where they have become integral parts of the everyday lives of large numbers of people, in the West these technologies remain mostly academic and speculative. While it is easy to dismiss Social Scoring as a concern for "those other people", it will inevitably be implemented in all places of the world in some form or another. And when we are presented with the initial proposals and gradually witness the erosion of our rights and freedoms, it will be much harder to fight. Who will be the first politician who demands that we all get credit scores? Who will be the first company to try to introduce scoring into an online environment? Who will be the first parent who demands that their child's school adopt social scorecards on their behalf?

2.1.1 United Kingdom: Live Facial Recognition & Health

In January 2020, the UK government revealed that they would be using Live Facial Recognition (LFR)[1] software on the streets of London in order to find suspects wanted by the police. The subjects in question are mainly criminals and terrorists, but even missing persons enters the pool. The statistics about reliability and accuracy are overall positive, taking into account the huge amount of crowd inside London and the few "dangerous" faces the system have to discriminate. The Metropolitan police (Met) said the system was 70% effective at spotting wanted suspects and falsely identified someone as wanted in one in a thousand cases, but Professor Pete Fussey[2] found it was verifiably accurate in just 19% of cases, so misclassification errors are consequentially prone to occur daily. The algorithm, once a person is scanned (1) if he/she is not present inside the database, his/her information is deleted in seconds, (2) otherwise an alert is sent to the officers near the identified suspect to ask him/her some questions.

It is believed that using facial recognition linked to databases of suspects is potentially the next big leap for law enforcement, as big as the introduction of fingerprints. The police have been working on it for years and the security services are also hugely interested. Nick Ephgrave, an assistant commissioner at the Met, said: "As a modern police force, I believe that we have a duty to use new technologies to keep people safe in London. Independent research has shown that the public support us in this regard". The only environments where the system is less effective are at night-time and when the crowds are so dense the people faces are only partially visible by the cameras. Many warnings have raised from the public, about the threat to human rights, privacy, non-discrimination, freedom of expression, association and peaceful assembly. On top of that the system is not open-source, so the inadequate transparency, non-existent public oversight and accountability measures are considered by many dangerous, oppressive and completely unjustified (*LFR in London*, 2020; *Social scoring systems: current state and potential future implications*, 2022).

In 2021, Boris Johnson tried to introduce a social credit system in order to fight obesity and improving the overall health condition among the British public. The scheme was intended to monitor family supermarket spending, rewarding those who reduced their calorie intake and buy more fruit and vegetables. The rewarding would have been given out as discounts or other incentives. The main advantage output of this system would have been the obesity and similar diseases rate reduction on the United Kingdom's health-care system (owned and operated by the state), alleviating the pressure and opening up the possibility of more beds, funding and doctors for other patients.

On the other hand, we have to take into consideration that one's body and one's health are, in a moral sense, just about the most private possession imaginable. For this reason, the state shouldn't have the right to handle and manipulate matters as invasive as one's blood pressure or one's body-fat percentages (*China's Social-Credit System Arrives on British Shores*, 2022).

[1] Live facial recognition. LFR is a real-time deployment of facial recognition technology, which compares a live camera feed (or multiple feeds.) of faces against a predetermined watchlist, in order to locate persons of interest by generating an alert when a possible match is found

[2] an expert on surveillance from Essex University who conducted the only independent review of the Met's public trials on behalf on the force.

2.1.2 Italy: Bologna's Smart Citizen Wallet

A new app bearing striking resemblance to the Chinese social credit system was presented during a press conference in Bologna. The "Smart Citizen Wallet" aims to reward people for "virtuous behavior", such as recycling, using public transport, managing energy well, and not getting fined. The app is scheduled for introduction in Bologna in fall 2022 while an experimental version is already active in Rome. The social credit system first made an entrance in China, where it's been evaluated since 2014. Privacy Network, an Italian tech company, issued a statement warning about the "legal, ethical, and social implications". The application of such a system in China allows a glimpse into future developments of such systems in Europe.

So far, the app is only aimed at rewarding good behavior. By improving their scores, citizens will gain points which then they will be able to spend on benefits such as discounts and free cultural activities. The exact nature of the benefits is currently being defined. But the possibility of marginalize who doesn't apply to the system, by punishing the subject by withdrawal of benefits is a present issue. The risk is to import and spread an Italian version copy of the social scoring made in China. The levity with which Bologna would like to overcome the "privacy problem" by ensuring that the system would work only with the consent of the citizen, who freely could download the relevant app made available by the municipality, is puzzling to say the least. We recall that in this regard that recital 43 of the GDPR (i.e., the EU Data Protection Regulation) expressly discourages the adoption of consent as a legal basis by public entities. Italy is all set to become the first European country to implement a social credit system that is government-sponsored (*Bologna Introduces Social Credit App to Promote "Virtuous Behavior"*, 2022; *Social Credit System in Europe*, 2022).

2.1.3 China: the social scoring world leader

In 2014, the Chinese government introduced the Plan for the Establishment of the Social Credit System (2014-2020) stating that they would promote the establishment of creditworthiness in four key areas, including:

1. Government performance
2. Credit score
3. Social score
4. Judicial credibility

The data is mainly obtained by the government departments, social organizations, etc. with the aid of facial recognition technology and big data analysis technology, aiming at "Regulating public behavior". The Xi Jinping regime's mass surveillance system has about 200 million public cameras in various places, and future estimates, when the social credit system is basically completed and fully operational, the number of public surveillance cameras in the Chinese territory is expected to reach 626 million. *Cinese Social Credit System, Wikipedia* (2022).

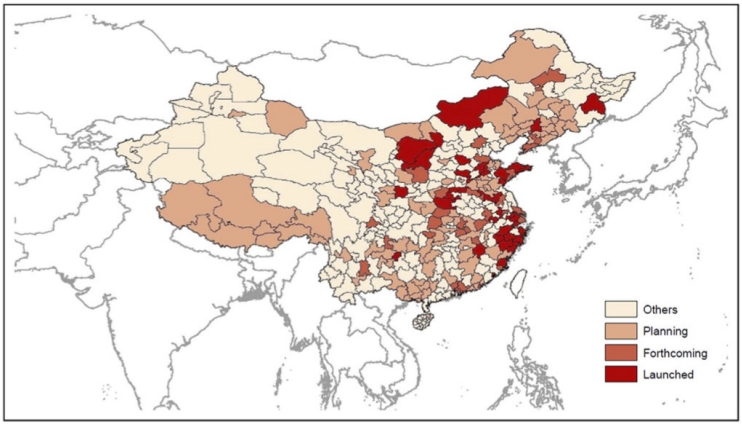
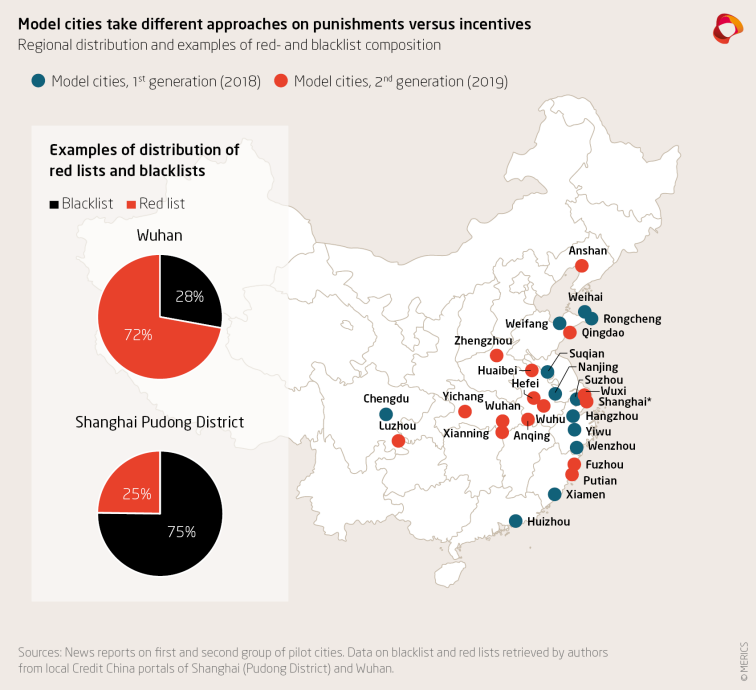


Figure 1: The geolocations of local SCSs in China, Liang and Chen (2022)

Incentives and punishments include the establishment of red lists and black lists: organizations or individuals with poor credit may face unfair strict supervision, and organizations with good credit will be able to obtain a "priority channel" for administrative approvals. For serious untrustworthy organizations or individuals, each government department is to adopt its own separate standards and put the organizations or individuals on their own blacklists. After being included in the blacklist, they will face "joint punishment" from multiple departments, and the blacklist is generally valid for 2-5 years, but if the untrustworthy person actively performs his obligations or takes some efforts to correct the incorrect behavior, the court can remove or reduce the duration of the punishment. As of June 2019, official data from the *National Development and Reform Commission* (1948) of the People's Republic of China shows that due to social score punishments 26.82 million people have been restricted

from purchasing air tickets, 5.96 million have been restricted from purchasing bullet train high-speed rail tickets, and 4.37 million "blacklisted" persons have voluntarily fulfilled their legal obligations.



In addition, some regions like Chongqing, Guangdong, Foshan, and Hebei restrict the children of people with a low score from enrolling in private schools, and some students are affected negatively from their parents' low score when applying for a university. This is a clear violation of various human rights, for example the equality principle and non-discrimination. We will explore more in depth the legal principles that are affected by Social Scoring in chapter 5.

Purpose Proponents believe that Chinese citizens are guided by the Social Scoring system towards paying more attention to honest and law-abiding social behavior and, as a consequence, improving the quality of life. Both national and municipal documentations claim that the decline of trustworthiness and morals can be solved via data and technologies. This vision indicates China's attempt to leverage digital technologies for addressing social and economic challenges.

On the other hand opponents, argue that the current policy undermines the legitimate rights and freedoms of individuals and units, especially the right to reputation, privacy, and personal dignity, and may become a tool for state surveillance and the Chinese Communist Party's suppression of dissent. To realize at which extent the Chinese authorities control their citizens, it is interesting to mention that private enterprises are not allowed to carry out personal credit assessment.

Although the sanctions are severe, in a research (*China's Social Credit System in 2021: From fragmentation towards integration* (2022)) it is found that only a small number of companies and individuals are affected. This is due to the fact that the main purpose is a general deterrence effect through the highly public nature of blacklists and extensive reporting about punishments in state media.

Process Since the complete pipeline of the Chinese SCSs is well beyond the scope of this paper, we are going to list its steps describing each at a high-level:

1. First, data warehouses involving the deployment of new infrastructures capable of gathering data from multiple sources are aggregated together. The types of data gathered are identity and behavior data.
2. Second, the heterogeneous qualities are transformed into a common metric. The analysis shows that the SCSs employ an assemblage of heterogeneous indicators to transform qualitative differences into quantifiable metrics for calculating how many points should be granted to a citizen.
3. Third, the behavioral data function as indicators to determine whether people earn or lose points. Indicators and scores hence express the degree to which people's past activities statistically relate to their creditworthiness.
4. In the final step, numerical scores are converted into ordinal levels of measurement, aiming to further categorize and rank citizens. Here two types of labels are used to label these rankings: (1) alphabetically-labeled ordinal measurement ranging from AAA to D and (2) name-labeled measurement (i.e., Excellent, Good, Fair, Poor, Risky).

2.2 Companies

The financial sphere used to be a place where humans needed their judgment to make qualitative decisions, but as companies grew and became more successful, they became increasingly rule-based and algorithmic. The same statement can be applied to human resources – or even to the interactions between people, which in many ways are much more regulated than any other part of life. It is hardly surprising that after so many years of getting used to scoring achieved by quantitative indicators, finally the system was found to be applied also to social relations. As mentioned multiple times through the paper, at the moment companies are not listed as subjects of the AI act, so theoretically they won't be limited in any way inside the European community.

2.2.1 Microsoft: quality and productivity scores

Microsoft Productivity Score is one of the more recent and controversial arrivals inside the Microsoft ecosystem. Aiming to provide visibility into employees' experience so the company can improve efficiency and empowerment, the score is sure to appeal to plenty of business leaders. Much misinformation ended up online, with disgruntled users complaining that the scoring system was akin to "workplace surveillance". According to a recent patent application, Microsoft is working on a "meeting insight computer system" that would track participants' body language, expressions, and other characteristics in order to give meetings a "quality score". The technology could be used for both in-person and online meetings, according to the filing. Additionally, Microsoft last month unveiled a "Productivity Score" that would have allowed businesses to keep track how their staff members are using Microsoft software. Today the Productivity Score tool focuses on:

- People experiences
  - **Communication**: How people share information through chat, email, community posts, and other tools for a unified employee experience;
  - **Content collaboration**: How people within the organisation create and collaborate on content in the cloud, wherever they are;
  - **Mobility**: How easy it is for team members to access and use files from the Microsoft 365 ecosystem to continue making progress on projects;
  - **Meetings**: How the organisation can leverage meeting best practices to improve meeting quality;
  - **Teamwork**: How the people in an organisation are using shared environments like Microsoft Teams and SharePoint for collaborative success;
- Technology experiences
  - **Endpoint analytics**: How organisations are impacted by performance and health issues with endpoint software and hardware. You can also receive recommendations for improvements;
  - **Network connectivity**: Visibility on the factors that may affect your network, combined with suggestions to address problematic areas;
  - **Microsoft 365 apps health**: Insights into the devices that have the best productivity and security components;
- Special reports
  - These are insights that help companies to measure specific changes in the workplace

The Microsoft Productivity Score offers fundamental insights into how the company uses various features and tools to function in a dynamic business environment. Moreover, the society may assess trends over time for things like technology adoption and compare its performance to some peer benchmarks offered by Microsoft, or obtain advice on how to improve. Microsoft gather data from all areas of the Microsoft 365 environment, to determine each score, which is displayed only to the Admin webpage. The rating is evaluated once the system has gathered data about technology and people experiences over a predetermined amount of time. The insights are mostly generalized even though the company can delve a little further into each of the indicators that Microsoft tracks. Microsoft Productivity Score doesn't ever provide with a detailed breakdown of what each employee in a company accomplishes on a daily basis. In general Microsoft Productivity Score isn't as invasive as many others, but rather a convenient way to get an insight into how people and technology are helping each business to evolve and discover which areas of experience need to focus on (*Microsoft Developing Workplace Surveillance System to 'Score' Meeting Productivity* (2020); *Understanding Microsoft Productivity Score* (2021)).

2.2.2 WeChat: introduction of the payment score

Every day, more than a billion individuals use WeChat to conduct a variety of tasks like reading, shopping, hailing taxis, renting umbrellas, and chatting. Furthermore, the Tencent app has covertly begun utilizing this kind of signal to decide if a user qualifies for benefits like deposit-free rental services. The scoring system, which the business refers to as the "WeChat Payments Score" in Chinese, was piloted on very few apps and soft-launched in eight cities last November. One of these is the power bank rental service Xiaodian, which is supported by Tencent and waives deposits for users after their points reach a particular threshold. It's simple to see how the incentives system could encourage users to explore WeChat's variety. WeChat's calculation of these points is unknown, but a TechCrunch test reveals that it takes into account a user's purchasing and his/her contract-fulfillment history.

WeChat's main rival in online payments, Ant Financial, an unit of Alibaba, has been using a similar evaluation engine named "Sesame Credit" since 2015. It measures a number of user data attributes, including buying behavior and contract fulfillment capacity, similar to WeChat's. For example an hotel may waive the deposit for guests with higher scores, providing an incentive to stay at the property. Through Ant's Alipay digital wallet, which reportedly has one billion users worldwide, you may purchase sesame points. The short-lived credit-rating program from Tencent is comparable to the WeChat payments score. In fact, the millions of people without bank records in China's developing financial system can be predicted as creditworthy using digital footprints. Due to this, Beijing hired internet firms like Tencent and Ant in 2015 to develop their own "social credit" scores as part of a pilot projects that was given governmental approval. Regulators started to take greater control of the entire credit-rating process as they grew increasingly concerned about the personal information that online lending companies were using. China has gone beyond most other nations, which primarily utilize credit ratings as a measure of financial legitimacy. China wanted everyone to be included in a national database by 2020 that includes not only financial history, but also social and moral background; this scheme has sparked worries about privacy and monitoring (*WeChat is quietly ranking user behavior to play catch-up with Alibaba* (2019)).

3 Social scoring and AI systems in EU

AI systems designed with social scoring in mind deserves special consideration since it can affect the access to and enjoyment of certain essential private and public services and benefits necessary for people to fully participate in society or to improve one's standard of living. In particular we will see the European Union (risk) approach with the *AI ACT* (2021) in the attempt to tackle problems that could emerge using AI, including Social Scoring in all its forms. Notice that at the current state restrictions only apply for public administrations. This means that every other private entity do not need to comply. Platform dedicated to dating or ridesharing for example, which inherently give a score to their users, are free to do so even if such a score could create an adverse impact for some users. Legislators are pondering if the AI ACT should be applied also to private companies, but we will discuss this topic better in chapter 5.

3.1 Prohibited practices

Social scoring is defined as the placing on the market, putting into service or use of AI systems by public authorities or on their behalf for evaluation or classification of the trustworthiness of natural persons (over a certain period) based on their social behavior or known predicted personal or personality characteristics, with the social score leading to either or both the following:

1. detrimental or unfavorable treatment of certain natural persons or whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected;
2. detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behavior or its gravity;

is prohibited by the *Article 5* of the *AI ACT* (2021).
This strict regulation would ban some products already on the market (e.g., *Bologna Introduces Social Credit App to Promote "Virtuous Behavior"* (2022) discussed in chapter 2) or for example China's SCS system would also be illegal and not deployable on the European market. However many subcategories of Social Scoring in a broader sense like Credit Scoring are complying with the AI ACT provided that they fall under the list of AI systems referred to in Annex III (*ANNEXES* (2021)). In the next sections we will focus on these high-risk AI systems which are regulated under *article 6*.

3.2 High-risk AI systems

In addition to the high-risk AI systems referred to in Section 1, AI systems referred to in Annex III shall also be considered high-risk. We only mention the points related to Social Scoring: the fourth point which covers employment, workers management and access to self-employment and the fifth point which covers access to and enjoyment of essential private services and public services and benefits. To summarize the following scope will be considered high-risk:

1. AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluation performance and behavior of person in such relationships;
2. AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;
3. AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use.

Here we will provide a brief list of the requirements for high-risk AI systems:

1. Compliance with the requirements (article 8)
2. Risk management system (article 9)
3. Data and data governance (article 10)
4. Technical documentation (article 11)
5. Record-keeping (article 12)
6. Transparency and provision of information to users (article 13)
7. Human oversight (article 14)
8. Accuracy, robustness and cybersecurity (article 15)

3.3 Credit score

The AI ACT states that "AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk AI systems, since they determine those persons' access to financial resources or essential services such as housing, electricity, and telecommunication services." The reason given is that an AI system used for this purpose may lead to discrimination of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation, or create new forms of discriminatory impacts. Nonetheless, on a broader scale (with respect to AI systems that may or may not include credit scoring) this restrictions should not stop the development and use of innovative approaches in the public administration, which would stand to benefit from a wider use of compliant and safe AI systems. When developing Credit Scoring AI systems the focus should be on ensuring the compliance with respect to *article 9* and *article 15*. In the former, paragraph 9 says that "for credit institutions regulated by Directive 2013/36/EU, the aspects described in paragraphs 1 to 8 shall be part of the risk management procedures established by those institutions pursuant to *article 74* of that Directive". In the latter it is trivial to understand why the importance of achieving an appropriate level of accuracy of the AI system, that must be resilient as regards errors, faults or inconsistencies. Moreover it is of paramount importance to make sure the system can withstand attacks such as:

- data breaches
- data poisoning (trying to manipulate training data)
- adversarial examples (input designed to cause the model to make a mistake)
- other exploits of the model due to intrinsic flaws

The problem is that while accuracy of the model is the top priority, cybersecurity is often underestimated as we can see from the Equifax[3] case. Bare in mind that Equifax collects and aggregates information on over 800 million individual consumers and more than 88 million businesses worldwide. In addition to credit and demographic data and services to business, Equifax sells credit monitoring and fraud prevention services directly to consumers. In 2017 it was affected by a data breach[4] resulting in the leak of the personal and banking data of over 150 million clients.

[3] Equifax Inc. is an American multinational consumer credit reporting agency headquartered in Atlanta, Georgia and is one of the three largest consumer credit reporting agencies, along with Experian and TransUnion (together known as the "Big Three").

[4] https://en.wikipedia.org/wiki/2017_Equifax_data_breach

4 Case studies in credit scoring

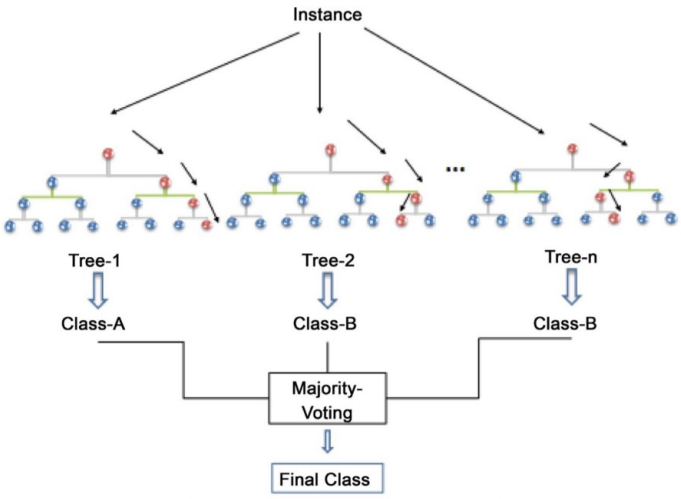
4.1 Deep Learning algorithms

The first paper we will analyze is titled "Wide and Deep Learning for Peer-to-Peer Lending"[5]. The goal of this paper is to propose a two-stage scoring approach to help lenders decide their fund allocations in peer-to-peer lending market.

Actually we are interested only in the first stage, where the *default prediction* is computed, while the second stage attempts to predict the profitability. This last step is irrelevant to us and we will focus only on the prediction about the ability of debtors to repay the loan or not (default). The algorithm is using "wide and deep learning" and achieves a 91% accuracy. Despite the fact that this is designed for P2P lending markets, similar algorithms could be implemented by banks in order to assess the ability of a client to repay a loan. Credit institutions using such AI system will however be subject to regulamentation. This particular case will lead to the blackbox[6] problem, so the provider should also produce some explainability documents in order to comply with the *article 13* dealing with transparency. So the algorithms would still be treated as high-risk regardless of the lack of clarity of the system.

4.2 Machine Learning algorithms

We can evaluate the results of algorithms using a more transparent approach on the paper "Combination of Random Forests and Neural Networks in Social Lending"[7]. For example, a simple linear function can achieve 65,6% accuracy while a random forest approach 73,3% (Fig. 4.2), and those are both Machine Learning approaches that do not pose the problem of transparency and so are considered as high-risk by the AI ACT. Indeed it is possible to see which features are the most determinant factors in the final resul: FICO is one of the main variables in explaining the output. A FICO score is a credit score created by the Fair Isaac Corporation (FICO). In America for example lenders use borrowers FICO scores along with other details on borrowers credit reports to assess credit risk and determine whether to extend credit.



Demonstration of the Random Forest methodology

To add another interesting example we cite the article "Can System Log Data Enhance the Performance of Credit Scoring?"[8]. The data used is the log data recorded by the mobile application system of KakaoBank, a leading internet bank used by more than 14 million people in Korea. The model described includes the independent variables derived from the system log data and the KCB[9] credit grade (instead of just the KCB credit grade already used as a form of indicator of creditworthness). The results show an improvement of 1.84% with respect to just using the KCB grade. Machine learning algorithms such the ones we presented are in compliance with the explainability requirements set by the AI ACT so the deployment on the market will be easier from a normative stand point. However the risk evaluation will be also high as for the AI systems affected by the blackbox problem.

[5] Bastani, Asgari, and Namavari (2018).

[6] We reffer to the blackbox problem when in system the steps that lead to the output are opaque and it is not possible to explain the reason of that result.

[7] Fu (2017).

[8] Kyeong, Kim, and Shin (2021).

[9] Korea Credit Bureau.

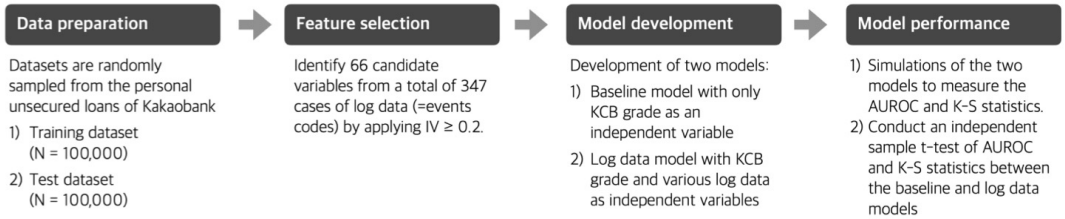


Figure 2: Schematic diagram of the model

5 Legal principles

Now that we have a rough idea of the state of the art for AI applications in Social Scoring and the approach that different countries are taking with respect to it, it is worth mentioning the different legal principles that are affected by AI. We will compare the relevant articles from *Human Declaration of Human Rights* (1948), the *Italian Constitution* (1947), the *European Convention on Human Rights* (1950) and the *Charter of fundamental rights of the European Union* (2000). From now on we will consider only the implication of AI system used by public administrations, but actually some of these issues may arise from private companies and platforms. However it is tricky to balance the right of freedom to conduct a business (e.g. Article 16 *Charter of fundamental rights of the European Union* (2000)) and the other principles that may conflict with it. This is not the point of this paper so we will not discuss it further, although it would be a worth mentioning digression.

5.1 Freedom and equality

It is trivial to see that every AI system giving a score to citizens can create inequality and discrimination if then the same score is utilized in order to grant or deny access to various kind of services, buildings or treatment. We have grouped together three similar principles that would be affected negatively by a Social Scoring system such the one adopted by China that we discussed earlier.

5.1.1 Equality principle

The most important one that would be voided if a social score is used in public administration would be the equality principle:

- All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood - Article 1 *Human Declaration of Human Rights* (1948)
- All citizens possess equal social dignity and are equal before the law, without distinction of sex, race, language, religion and political orientation, personal and social conditions. It is the duty of the Republic to remove economic and social obstacles which, by limiting the freedom and equality of citizens, prevent the full development of the natural person and the actual participation of all workers in the political, economic and social organisation of the country. - Article 3 *Italian Constitution* (1947)

5.1.2 Prohibition of discrimination

- All are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination. - Article 7 *Human Declaration of Human Rights* (1948)
- The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status. - Article 14 *European Convention on Human Rights* (1950)
- Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited. - Article 21 *Charter of fundamental rights of the European Union* (2000)

5.1.3 Freedom of expression

While equality and discrimination would represent a direct effect of a "barrier to entry" enforcement (based on the social score), an indirect negative effect that it is worth noting is the fear of expressing an opinion in order not to lower the score.

- Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers. - Article 19 *Human Declaration of Human Rights* (1948)
- Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises. - Article 18 *European Convention on Human Rights* (1950)
- Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. - Article 11 *Charter of fundamental rights of the European Union* (2000)
- Everyone has the right to freely express their ideas through speech, in writing and by any other means of communication... - Article 21 *Italian Constitution* (1947)

5.2 Right to medical care

Another right that should be pointed out is the right to medical care. For example, in China people with a low social score would receive basic treatments in case of emergency but nothing more. This is a controversial decision, because actually implementing a social score in healthcare would violate:

- Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control. - Article 25 *Human Declaration of Human Rights* (1948)
- Everyone has the right of access to preventive health care and the right to benefit from medical treatment under the conditions established by national laws and practices. A high level of human health protection shall be ensured in the definition and implementation of all Union policies and activities. - Article 35 *European Convention on Human Rights* (1950)
- The Republic shall safeguard health as a fundamental right of the individual and as a social interest and shall guarantee free medical care to the indigent. - Article 32 *Italian Constitution* (1947)

5.3 Right to privacy

Lastly, but not less important, we have right to privacy. As students of the Artificial Intelligence System master degree, we are aware of the critical role that data has in training AI systems. But as a society we underestimate the amount of information we share and the magnitude of power that is in the hand of whom control such large datasets. "With great power comes great responsibility", the responsibility to protect the data from bad actors or competitors, because unfortunately data breaches occur daily even to big corporations such as Google or Microsoft. This to say that even if we trust the entity which uses our information, it is in everyone interest to protect the principle of privacy.

- No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks. - Article 12 *Human Declaration of Human Rights* (1948)
- Everyone has the right to respect for his private and family life, his home and his correspondence. - Article 8 *European Convention on Human Rights* (1950)
- Everyone has the right to respect for his or her private and family life, home and communications. - Article 7 *Charter of fundamental rights of the European Union* (2000)
- Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority. - Article 8 *Charter of fundamental rights of the European Union* (2000)

5.4 Principle of good administration

In order to analyze fairly and without biases an implementation of Social Scoring in our society, we decided to also list this potential improvement out of all the negative sides we already mentioned. Given the ability of a machine to process a huge amount of decisions, a properly designed AI system could be many orders of magnitude faster and efficient compared to a human in tasks that are carried out by public administrations around the world. One of such principles that is affected positively by this technology is the article 97 of the *Italian Constitution* (1947):

- Public administrative offices shall be organised according to the provisions of law, so as to ensure the efficiency and impartiality of the administration. The regulations of the offices shall lay down the areas of competence, the duties and the responsibilities of the officials. Appointments in public administration shall be made through competitive examinations, unless otherwise set forth by law.

6 Conclusion

Within our research we attempted to analyze the main concept of Social Scoring, identifying its fields of application linked with the Artificial Intelligence, the purpose with which such systems were created, and the effects they could produce within the society in which we live, both in the ethical and legislative fields. We attempted to expose the current state of the art of such systems, delving into how they are differently understood in the European and non-European contexts, focusing in particular on the Chinese case history.

In analyzing these aspects, it became necessary to compare them with our reality as Italian and European citizens. The contrast that arises between a need for the development of AI (which, as students in this field, we feel obliged to promote within limits) and respect not only for personal privacy, but also for all the legal principles mentioned in chapter 5 appeared immediately. And these are not the only balances to be considered: we also assessed the imbalance governing applications of Social Scoring systems between public and private authorities as potentially risky and counter-intuitive. Public authorities, in fact, are required to comply with a whole series of limitations that we previously described within our research. In contrast, private entities do not have to be subject to them, perhaps following an ineluctable power and, above all, not in line with the standards of transparency that any AI system should respect. "With great power comes great responsibility", someone said, and we are not sure we can say that private authorities would be able to handle these responsibilities.

Another food for thought we would like to mention concerns what was also mentioned in the opening paragraph, about the utopian goal of these Social Scoring systems. In our opinion, the law cannot be considered a sufficient aspect to change the intrinsic nature of citizens. In particular, we disagree with the design of such systems the moment they turn out to be based on the "theory of punishment". This theory, in our opinion turns out to be ineffective, especially because it is based on principles of revenge against a less than virtuous citizen that risk eroding the fundamental rights guaranteed to any citizen. Rights from which we wish never to be taken away. These are not Social Scoring systems that we hope will be introduced within our communities and into our daily lives.

In conclusion, therefore, we feel like making a proposal for a possible solution of using Social Scoring systems. In our opinion, these tools could be promoted as a method solely of rewarding rather than penalizing. Starting from the fact that every fundamental right must be guaranteed to any citizen, these systems could become an integral part of a community aiming for continuous improvement if they provided solely for the concept of merit. This could be implemented by making sure that those citizens who prove to be most virtuous can obtain benefits (at the fiscal level, for example) and it could prove to be a win-win compromise in all aspects and, in our opinion, would not promote any policy of discrimination: everyone, in fact, turns out to be free to choose according to his or her own opinion whether to aim to obtain certain benefits, without having to fear sanctions and penalties if he or she decides not to follow that path. Our reasoning is also concerned with the extent of these benefits to be awarded: they should always represent an award of not excessive weight, so that they do not become a discriminatory factor. An imbalance too great between deserving and undeserving citizens would lead to the creation of too diverse treatment, with the risk of creating classes and separations over time. Instead, the use of small and calibrated benefits could lead to the desired goal.

We believe that this could be a proper use of Social Scoring systems, applicable to our daily reality, which could, over time, be positively affected by a drive for self-improvement promoted by Artificial Intelligence.

References

Ai act. (2021). https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC$_1 format = PDF$.

Annexes. (2021). https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC$_1 format = PDF$.

Bastani, K., Asgari, E., & Namavari, H. (2018, 10). *Wide and deep learning for peer-to-peer lending.*

Bologna introduces social credit app to promote "virtuous behavior". (2022). https://europeanconservative.com/articles/news/bologna-introduces-social-credit-app-to-promote-virtuous-behavior/.

Charter of fundamental rights of the european union. (2000). https://www.europarl.europa.eu/charter/pdf/text$_e n.pdf$.

CHENEY-LIPPOLD, J. (2017). *We are data: Algorithms and the making of our digital selves.* NYU Press. Retrieved 2022-11-22, from http://www.jstor.org/stable/j.ctt1gk0941

China's social-credit system arrives on british shores. (2022). https://www.theguardian.com/technology/2020/jan/24/met-police-begin-using-live-facial-recognition-cameras.

China's social credit system in 2021: From fragmentation towards integration. (2022). https://merics.org/en/report/chinas-social-credit-system-2021-fragmentation-towards-integration.

Cinese social credit system, wikipedia. (2022). https://zh.m.wikipedia.org/zh-hans/.

European convention on human rights. (1950). https://www.echr.coe.int/Documents/Convention$_E NG.pdf$.

Fu, Y. (2017). Combination of random forests and neural networks in social lending. DOI: 10.4236/jfrm.2017.64030

Heikkila, M. (2021). Europe eyes strict rules for artificial intelligence. *POLITICO Europe*.

Human declaration of human rights. (1948). https://www.un.org/sites/un2.un.org/files/2021/03/udhr.pdf.

Italian constitution. (1947). https://www.quirinale.it/allegati$_s tatici/costituzione/costituzione_i nglese.pdf$.

Kyeong, S., Kim, D., & Shin, J. (2021, 12). Can system log data enhance the performance of credit scoring?—evidence from an internet bank in korea. *Sustainability*, *14*, 130. DOI: 10.3390/su14010130

Lfr in london. (2020). https://www.theguardian.com/technology/2020/jan/24/met-police-begin-using-live-facial-recognition-cameras.

Liang, F., & Chen, Y. (2022). The making of "good" citizens: China's social credit systems and infrastructures of social quantification. *Policy & Internet*, *14*(1), 114–135.

Maamar, N. (2018). Social scoring. *Computer und Recht*, *34*(12), 820–828. Retrieved 2022-11-22, from https://doi.org/10.9785/cr-2018-341212 DOI: doi:10.9785/cr-2018-341212

Microsoft developing workplace surveillance system to 'score' meeting productivity. (2020). https://epic.org/microsoft-developing-workplace-surveillance-system-to-score-meeting-productivity/.

National development and reform commission. (1948). https://en.ndrc.gov.cn/.

Social credit system in europe. (2022). https://www.nspirement.com/2022/05/21/social-credit-system-in-europe.html.

Social scoring systems: current state and potential future implications. (2022). https://www.kaspersky.com/blog/social-scoring-systems/.

Sánchez-Monedero, J., & Dencik, L. (2022). The politics of deceptive borders: 'biomarkers of deceit' and the case of iborderctrl. *Information, Communication & Society*, *25*(3), 413-430. Retrieved from https://doi.org/10.1080/1369118X.2020.1792530 DOI: 10.1080/1369118X.2020.1792530

Understanding microsoft productivity score. (2021). https://www.uctoday.com/collaboration/understanding-microsoft-productivity-score/.

Wechat is quietly ranking user behavior to play catch-up with alibaba. (2019). https://techcrunch.com/2019/01/14/wechat-credit-scoring/.