Social Credit System

* The premise behind a nationwide social rating system itself is relatively simple: every citizen receives a certain score to start, and certain actions either lower or increase your score. For example, donating to charity would increase your score, while buying cigarettes would lower it.
* People can then either be rewarded or punished based on their rating.
* The government, for example, could restrict a person’s travel, prevent them from entering the best universities, or even take their dog away if their score drops low enough.
* Two of the biggest unanswered questions are what the system will actually look like in practice and how the system operates from a technological standpoint. There are several possible reasons for this.
* First, the developers of the system understand that the more they reveal about the system, the more they open it up to vulnerabilities (more on that later).
* Second, it’s entirely possible that the creators themselves do not fully understand how this system works. Many rating systems used today are built on large volumes of historical data and machine learning models that predict the future behavior and successes of system participants. For example, Microsoft uses such models to rank the skill level of players in online games, banks evaluate the reliability of potential borrowers when they submit applications for loans, and several companies have even tried to automate the process of reviewing resumes for open vacancies. In these situations, developers put their trust in the algorithms.
* AI is based on machine learning algorithms, which, despite being generally lauded by the technological industry as a “cure-all”, are far from perfect. Like any other computer system, they can be prone to errors. For example, in cybersecurity, machine-learning algorithms are used to rapidly detect previously unknown malware. However, there is a problem: the higher the detection rate, the higher the chance you’ll run into “false positives“—i.e., the system determines a non-malicious file is malicious. This happens due to the very nature of how machine learning works: the ML-based system doesn’t look into the details of an object but compares the way it “looks” to other known objects. In some cases, “clean” objects may “look” a lot like malicious ones and a system that makes decisions based on scoring would most likely confirm the object as a malicious one. When applied to a world where people’s behavior is evaluated by a fully automated system, this particularity of machine learning systems may lead to multiple unpleasant situations where an innocent person is confirmed as the one behind “wrong” actions.
* These systems are also susceptible to issues such as developer bias, false correlations, and feedback loops, and, unless specifically included by the developer, the algorithms do not factor in ethical considerations. To simply input massive quantities of information into a machine learning system and then accept the result without any critical assessment could lead to a host of unintended consequences, including choices that ultimately infringe upon the rights of certain citizens.
* Lastly, should the control over a particular system be concentrated in the hands of only one social group, their ability to change the rules of how the system works may significantly influence the life of those social groups who are not in a position to influence the scoring rules.