Machine Learning models tend to reproduce and amplify biases. Some techniques, relying on Fairness, allow to fight against these bias. Fairness is one of the sub-topic of Trustable Machine Learning. Another sub-topic is explainability. In Machine Learning, we use some techniques often very strongly non-linear and understanding models behavior is a hard task. In this talk, we will use an adversarial network coming from literature to limit influence of some parameters on a classifier to achieve Fairness. We will measure the gain of Fairness compared to a classifier with the base rate metric named p-rule, with metrics based on the distance between two distributions and with the Differential Fairness. To have local explanation of the difference between the classical and the fair classifier, we will compute the Shapley Value, a classical explanation technique coming from Game Theory, whose definition will depend on a reference population, and not the average prediction as often seen. Finally, we will show the contribution in terms of data protection of the fair classifier. This work is funded under the SPARTA project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830892.