Punishing Autonomous Agents.

By - Yohannis Kifle Telila

The involvement of Autonomous agents in our society has become more and more apparent and are being used in different sectors. From self-driving cars to autonomous robots performing medical surgeries without the involvement of human professionals, the involvement of these agents in our society are wide. Under some circumstances, these agents could engage in activities that are criminal or against the law. For instance, autonomous driving cars disobeying stop signs or even causing serious injuries against pedestrians. These transformative developments of autonomous agents in our society should also be supported by regulatory implications that address the issue of responsibility and accountability in the case where these agents are involved in criminal activities.

The current criminal law fails to address who to hold accountable in the case of criminal activities. In traditional law, the driver or any human will become liable for irresponsible actions that lead to an accident and face necessary legislative sanctions. This liability is not easy when we come to autonomous agents. Determining who is at fault if an autonomous agent made a wrong decision and as a consequence acts against the law is not easy under the current criminal law. Who to hold accountable and liable for the wrongdoing of the agent could depend on how we conceive these autonomous systems.

What differentiates humans from autonomous agents or artificial intelligence enhanced machines is the fact that humans are morally responsible for their actions and the punishments that are placed for the wrongdoing of a person would not work in case of autonomous agents. For instance, a person could face jail time for a specific time or could be asked to pay a fine for his or her wrong actions. Sentencing an autonomous agent to jail time or making them pay a fine will not make sense because the autonomous agent does not have any mental state or a moral and the punishments or liability on these agents will not achieve the main objective of punishments. If a human committed a crime, we would assume that person will not make the same mistake again and we hope that person will learn from his mistake and will be a better person after finishing his/her sentence. This is not the same for the autonomous agents.

As long as these agents are not fully independent the liability and the punishment should be distributed among all the stakeholders involved in the development of these agents. Including the programmers involved in the development of a software, a manufacturer or a director of the company leading the development of these autonomous agents should be the ones responsible for the wrongdoing of these agents as long as these agents are not fully independent and morally responsible entities.

In conclusion, applying the current criminal law to electronic agents will not be as effective as it is for humans. So the liability for the wrongdoing of these agents should be distributed among the makers of the agents.