

Abstract: The development of increasingly autonomous robotic systems threatens to undermine the foundational principle that a human agent (either individual or corporate) is responsible, and potentially accountable and liable, for the harms caused by the deployment of any technology. "Many hands" participate in the building of complex information systems. Designers and engineers cannot always predict how complex systems will act in new situations with new inputs. And yet, robotics is perceived as a transformative technology whose development should be stimulated and whose benefits will outweigh the risks. There have been, and will continue to be, calls to lower liability to stimulate the development of the robotics industry, to provide 'no fault' insurance to facilitate the introduction of robots used in the transportation and service industries, and to develop robots for military applications, e.g., drones, and then introduce those robots for domestic applications with little or no public debate.

Rather than dilute the principle of a responsible human agent, that principle should be reinforced for the design and deployment of new robotic systems. This reinforcement will have a number of elements:

- 1.) Engineering for responsibility: Considerations over who will be responsible for the actions of robots, or how responsibility will be apportioned if harms occur, should be made an integral part of the design process. This will turn engineers away from design solutions in which the attribution of responsibility could be impossible.
- 2.) Machines must not make decisions that are harmful to humans: A strong international principle that machines should not be making life or death decisions or independently initiate harms to humans must be established. Once such a principle is in place we can go on to the more exacting discussion as to the situations in which robots and information systems are indeed an extension of human will and intention and when their actions are beyond direct human control.
- 3.) Limit any dilution of responsibility: Any dilution of responsibility for the actions of robots must be short-term and not lead to undermining the core principle that a human agent (individual or group) is responsible for all actions taken by 'intelligent' machines.

4.) Oversight of developments in robotics: A governance mechanism should be put in place to monitor whether the lines of responsibility have been established for new systems being deployed.