The Robot As Enemy?

Isaac Asimov

It was back in 1942 that I invented “the Three Laws of Robotics,” and of these, the First Law is, of course, the most important. It goes as follows: “ A robot may not injure a human being, or, through inaction, allow a human being to come to harm.” In my stories, I always make it clear that the Laws, especially the First Law, are an inalienable part of all robots and that robots cannot and do not disobey them.

I also make it clear, though perhaps not as forcefully, that these Laws aren’t inherent in robots. The ores and raw chemicals of which robots are formed do not already contain the Laws. The Laws are there only because they are deliberately added to the design of the robotic brain, that is, to the computers that control and direct robotic action. Robots can fail to possess the Laws, either because they are too simple and crude to be given behavior patterns sufficiently complex to obey them or because the people designing the robots deliberately choose not to include the Laws in their computerized makeup.

So far—and perhaps it will be so for a considerable time to come—it is the first of these alternatives that holds sway. Robots are simply too crude and primitive to be able to foresee that an act of theirs will harm a human being and to adjust their behavior to avoid that act. They are, so far, only computerized levers capable of a few types of rote behavior, and they are unable to step beyond the very narrow limits of their instructions. As a result, robots have already killed human beings, just as enormous numbers of noncomputerized machines have. It is deplorable but understandable, and we can suppose that as robots are developed with more elaborate sense perceptions and with the capability of more flexible responses, there will be an increasing likelihood of building safety factors into them that will be the equivalent of the Three Laws.

But what about the second alternative? Will human beings deliberately build robots without the Laws? I’m afraid that is a distinct possibility. People are already talking about security Robots. There could be robot guards patrolling the grounds of a building or even its hallways. The function of these robots could be to challenge any person entering the grounds or the building. Presumably, persons who belonged there, or who were invited there, would be carrying (or would be given) some card or other form of identification that would be recognized by the robot, who would then let them pass. In our security-conscious times, this might even seem a good thing. It would cut down on vandalism and terrorism and it would, after all, only be fulfilling the function of a trained guard dog.

But security breeds the desire for more security. Once a robot became capable of stopping an intruder, it might not be enough for it merely to sound an alarm. It would be tempting to endow the robot with the capability of ejecting the intruder, even if it would do injury in the process—just as a dog might injure you in going for your leg or throat. What would happen, though, when the chairman of the board found he had left his identifying card in his other pants and was too upset to leave the building fast enough to suit the robot? Or what if a child wandered into the building without the proper clearance? I suspect that if the robot roughed up the wrong person, there would be an immediate clamor to prevent a repetition of the error.

To go to a further extreme, there is talk of robot weapons: computerized planes, tanks, artillery, and so on, that would stalk the enemy relentlessly, with superhuman senses and stamina. It might be argued that this would be a way of sparing human beings. We could stay comfortably at home and let our intelligent machines do the fighting for us. If some of them were destroyed—well, they are only machines. This approach to warfare would be particularly useful if we had such machines and the enemy didn’t.

But even so, could we be sure that our machines could always tell an enemy from a friend? Even when all our weapons are controlled by human hands and human brains, there is the problem of “friendly fire. “ American weapons can accidentally kill American soldiers or civilians and have actually done so in the past. This is human error, but nevertheless it’s hard to take. But what if our robot weapons were to accidentally engage in “friendly fire” and wipe out American people, or even just American property? That would be far harder to take (especially if the enemy had worked out stratagems to confuse our robots and encourage them to hit our own side). No, I feel confident that attempts to use robots without safeguards won’t work and that, in the end, we will come round to the Three Laws.