

A brief note on “Robots, Research, and Responsibility”

by Hans-Georg Stork

Robots have come a long way since the Czech writer Karel Čapek used this term some ninety years ago to denote rather frightening creatures not unlike Golems or Frankenstein’s monster, yet workers all the same. Today “*robot*” mainly refers to electro-mechanical devices of all shapes and sizes, designed and built to help people do jobs that are physically strenuous, potentially dangerous, repetitive and tiring, or simply impossible to do without suitable technical support. They can be stationary or mobile, and handle and/or transport physical objects, large or small, heavy or light, depending on the kind of service they are supposed to deliver. The history of industrialisation could well partly be written from the perspective of progress in building these time-and-motion saving devices, and more and more complex and powerful tools of this kind.

Refining and improving the mechanics of robots and their sensorial capacities (including ones that living organisms do not possess) has always been a major concern of engineers. Reducing the amount of human intervention in the operation of these machines, has been another persistent trend, leading for instance to numerically controlled (NC) machine-tools. Ultimately, however, this means more than merely automating the completion of a task according to some preset rules. It means that within certain limits, machines ought to be able to take decisions autonomously, independently of external (e.g., remote) control, on how to proceed with a given task, should new conditions arise unexpectedly. This could be a roving robot that is supposed to retrieve some object from a distant place but on its way encounters an unexpected obstacle.

Ease of use, safety and partial autonomy are essential if robotic devices are to leave the shop floor and strictly controlled environments, and become useful helpers and companions for people, including those with special needs. Just think of the potential advantages if the machines we live with could, of their own accord, bare some of the human mental efforts needed to carry out more or less demanding tasks - apart from manufacturing all sorts of goods. This could include steering a wheelchair, driving a car, guiding a blind person, performing precision surgery, operating a leg amputee’s prosthesis or many of our daily chores.

None of these machines is expected to solve chess conundrums or any other classical Artificial Intelligence problem; they are not supposed to keep track of budgets or stock. But they should have their wits about them, if for instance, they might need to recognise a certain object if viewed from a different angle or under different lighting conditions. Other systems will need to understand their users’ intentions and what they are saying in plain natural language. All of them would have to understand to a greater or lesser degree, aspects and features of their environment. We may for instance want robots to know or be able to learn what they can do with certain objects in their worlds: what the handle of a mug is good for, a dish washer, the curb of the pavement along a busy street. . .

Machines and systems which in this sense are *cognitive* do not necessarily need to be as intelligent as humans, or to be as conscious as humans of what they are doing. But they are supposed to have some of the same capabilities as animals, who have much less grey matter at their disposal than we humans. Engineers still have a lot to learn from solutions that natural evolution has developed over billions of years.

Considerable research effort taking new, multidisciplinary approaches is needed to significantly advance the engineering of the machines and systems described. It adds to our knowledge and prowess in building robotic systems which are ever safer, more robust, efficient, easy to use, and - where needed - more autonomous. It will potentially be re-

warded with greatly increased productivity of human labour and the creation of new useful products and services.

But apart from the many scientific, technical and socio-economic challenges equally challenging *ethical questions* remain. They are of particular interest and importance if public funding is involved. We can not and must not curb scientific curiosity but we should ask: Are there general principles that may guide public funding of research and the use of its results – beyond innovation and competitiveness?

Bertolt Brecht, in his “*The Life of Galilei*”, had the great scientist say: “*I maintain that the only goal of science is to alleviate the drudgery of human life.*” Sound advice, indeed! We should fund research whose results could help create better living conditions for everyone on this planet, and research that helps us to better understand ourselves and the world we live in. The former is not possible without the latter.

Clearly, there are at least two faces to whatever we discover or invent. We are all too aware of the shady sides of nuclear technology, biotechnology, and of the technologies which help to speed up the process of global warming. All of us - decision makers, researchers and citizens alike - should ensure that research results can not and will not be used to increase the suffering of people or harm our planet. Past experience shows that this is not easy. Free movement of researchers, scientific knowledge and technology is certainly a part of achieving this. Ultimately, however, legislators need to set the rules, with due regard to the democratic process.

Cognitive Systems and Robotics research may well be in line with Galilei’s advice. A welcome side-effect could even be the improvement of our understanding of the human condition. Yet we must be aware of potential hazards and downsides. The prospect of being cared for by a robot when we are old and frail may be frightening, comforting, or amusing, depending on one’s point of view. But would we really choose to replace a human carer by a machine, if such an option were available? Or should we not use these machines to complete other tasks or roles and give people more time to care for and help each other? The choice is ours.

A more serious concern is linked to the concept of an autonomous machine. This could be a self-controlling road vehicle which may become a reality rather sooner than later given the speed of current technological advancement. It could be a robot designed to replace a soldier in the battlefield. Who is responsible for its actions? Who is liable? Can such a machine behave *ethically* of its own accord?

I believe the answer is a firm “no”. Machines should be designed, built, programmed, trained so that they can render useful services. While possibly requiring limited autonomy they must for example never be allowed to kill. In the last instance, they must always be controlled by people. Machines – no matter how sophisticated - are as “ethical” as the people who design, build, program, train and use them.

We must not forget that man-made machines are categorically different from natural living, feeling, and thinking beings. The more we fancy machines to be human-like the higher the risk of us becoming machine-like ourselves. The more we rely on machines to make decisions that only we can justifiably make, the more we deprive ourselves of our authority, independence and our essential human characteristics. And contrary to nightmarish musings about robots taking over the planet and about robot-rights: man-made machines – no matter how sophisticated - have no rights and should not be feared; we can switch them off, take them off line or, ultimately, dismantle them.

We, jointly and individually, have to take full responsibility for what we are doing, good or bad, constructive or destructive, through our own inventions and creations, to each other and our world at large.