BRUNO SICILIANO - GUGLIELMO TAMBURRINI

ETHICS AND ROBOTICS IN THE FOURTH INDUSTRIAL REVOLUTION

1. *Introduction* 2. *Robots and robotics* 3. *Field and service robots*
4. *Human-robot interaction* 5. *Robotics for Industry 4.0*
6. *Roboethics and human-robot interaction in the workplace*
7. *Increasing autonomy of surgical robots and human responsibility*
8. *Roboethics and technological unemployment*

ABSTRACT: *ETHICS AND ROBOTICS IN THE FOURTH INDUSTRIAL REVOLUTION*
The current industrial revolution, characterised by a pervasive spread of technologies and robotic systems, also brings with it an economic, social, cultural and anthropological revolution. Work spaces will be reshaped over time, giving rise to new challenges for human-machine interaction. Robotics is hereby inserted in a working context in which robotic systems and cooperation with humans call into question the principles of human responsibility, distributive justice and dignity of work. In particular, the responsibilities for using a robotic system in a surgical context will be discussed, along with possible problems of medium- or long-term technological unemployment to be tackled on the basis of shared concepts of distributive justice. Finally, the multiple dimensions of human dignity in the working context are dealt with in terms of dignity of work, dignity at work and dignity in human-machine interaction.

1. *Introduction*

Robotics is a young and evolving science. For the Industry 4.0 programme it is the first of the enabling technologies that, as a whole, are considered essential to foster growth and employment. According to the definition of the European Commission, the enabling technologies are «knowledge-intensive and associated with

high R&D activity, rapid innovation cycles, substantial investment expenditure and highly skilled jobs»; hence, the systemic relevance potential of robotics and other technologies, as capable to feed the value chain of the production system with a capacity to innovate processes, products and services in all economic sectors of human activity. In the midst of the fourth industrial revolution, the key words of robotics are collaboration and autonomy. In the field of traditional industrial automation systems, robots were built and used to perform repetitive operations with high precision and speed. However, they were confined for safety reasons to spaces far from humans. In the new generation factories, humans are flanked or replaced by collaborative robots, capable of working together with the worker in a safe and reliable manner, and by autonomous robots, capable of moving and working even in the presence of uncertainty and variability in the environment.

Today and in the future, the objective of advanced robotics research is to flesh out artificial intelligence by creating automata in which physical and cognitive skills converge for the support of the elderly or the disabled, to reduce execution time and improve productivity of workers on production lines, to minimise the environmental impact of people and goods transportation, and to promote the progress of diagnostic and surgical techniques. The current industrial revolution, with all its pervasive dimension in terms of technologies and robotic systems, is also an economic, social, cultural and anthropological revolution. Work spaces will be reshaped over time, giving rise to new challenges for human–machine interaction.

This is where roboethics comes into play, in a context in which robotic systems and interaction with humans call into question the principles of human responsibility, distributive justice and dignity of work. In view of the constraints or objectives worthy of moral consideration to be placed on technological development,

a thorough ethical reflection is needed, focusing on the development of systems with growing autonomy in harmony with the moral autonomy and the attending responsibilities of human beings. Medium – or long-term technological unemployment – a time-honoured subject of investigation in economics and ethical reflection since the first industrial revolution – is another issue that will be discussed here in connection with the ethical implications of robotics, and its possible impact on the loss of certain types of jobs and the creation of new ones. Indeed, a reflection is needed in the light of distributive justice principles to assess whether there is a social duty to compensate for job losses, should pessimistic views concerning unemployment effects of robotic innovation on human employment come true. The technological unemployment issue is part of broader ethical discussions about robotics and work, which concern multiple dimensions of human dignity: dignity of work, dignity at work and dignity in human–machine interaction.

2. *Robots and robotics*

Robotics has profound cultural roots. Over the course of centuries, human beings have constantly attempted to seek substitutes that would be able to mimic their behaviour in the various instances of interaction with the surrounding environment. Several motivations have inspired this continuous search referring to philosophical, economic, social and scientific principles.

Asimov's books and science fiction have undoubtedly influenced the man and the woman in the street that continue to imagine the robot as an android who can speak, walk, see, and hear. In reality, the robot (derived from the term *robota* that means executive labour in Slav languages) is defined as any machine that is able to carry out tasks in an automatic way to replace or improve human work.

In order to understand the technical meaning of the term robot, we may refer to the definition of robotics as the intelligent

connection between perception and action[1]. With reference to this definition, the action of a robotic system is entrusted to a locomotion apparatus to move in the environment (wheels, crawlers, legs, propellers) and/or to a manipulation apparatus to operate on objects present in the environment (arms, end effectors, artificial hands), where suitable actuators animate the mechanical components of the robot. The perception is extracted from the sensors providing information on state of the robot (position and speed) and its surrounding environment (force and tactile, range and vision). The intelligent connection is entrusted to a programming, planning and control architecture which relies on the perception and available models of the robot and environment and exploits learning and skill acquisition.

Robots started to become widely used in industry since the 1970's. The main factors having determined the spread of robotics technology in an increasingly wider range of applications in the manufacturing industry, especially in the automobile industry, are reduction of manufacturing costs, increase of productivity, improvement of product quality standards and, last but not least, the possibility of eliminating harmful or off-putting tasks for the human operator in a manufacturing system. Industrial robotics is to be considered as a well-assessed technology by now.

On the other hand, with the term advanced robotics we usually refer to the science studying robots with marked characteristics of autonomy, operating in scarcely structured or unstructured environments, whose geometrical or physical characteristics would not be known a priori. Nowadays, advanced robotics is still in its youth. It has indeed featured the realisation of prototypes only, because the associated technology is not yet mature. There are many motivations which strongly encourage advances in knowledge within this field. They range from the need for automata whenever

[1] Cfr. B. Siciliano, L. Sciavicco, L. Villani, G. Oriolo, *Robotics: Modelling, Planning and Control*, 2nd edition, Springer, Berlin, Heidelberg 2009.

human operators are not available or are not safe (field robots)[2], to the opportunity of developing products for potentially wide markets which are aimed at improving quality of life (service robots)[3].

3. *Field and service robots*

In field applications, robots are deployed in areas where human beings could not survive or be exposed to unsustainable risks. Such robots should carry out exploration tasks and report useful data on the environment to a remote operator, using suitable onboard sensors. Typical scenarios are the exploration of a volcano, the intervention in areas contaminated by poisonous gas or radiation, or the exploration of the deep ocean or space. As is well known, NASA succeeded in delivering some mobile robots (rovers) to Mars which navigated on the Martian soil, across rocks, hills and crevasses. Such rovers were partially teleoperated from earth and have successfully explored the environment with sufficient autonomy. Some mini-robots were deployed on September 11, 2001 at Ground Zero after the collapse of the Twin Towers in New York, to penetrate the debris in the search for survivors.

A similar scenario is that of disasters caused by fires in tunnels or earthquakes; in such occurrences, there is a danger of further explosions, escape of harmful gases or collapse, and thus human rescue teams may cooperate with robot rescue teams. Also in the military field, unmanned autonomous aircrafts and missiles are utilised, as well as teleoperated robots with onboard cameras to explore buildings.

Autonomous vehicles are also employed for civil applications, i.e., for mass transit systems, thus contributing to the reduction of pollution levels. Such vehicles are part of the so-called Intelligent Transportation Systems (ITS) devoted to traffic management in urban areas. Another feasible application where the adoption of mobile robots offers potential advantages is museum guided tours.

Many countries are investing in establishing the new market of service robots which will co-habitat with human beings in everyday life.

[2] Cfr. B. Siciliano, O. Khatib, *Springer Handbook of Robotics*, 2nd edition, Springer, Berlin, Heidelberg 2016 (Part F).
[3] *Ibid.*, Part G.

Technology is ready to transform into commercial products the prototypes of robotic aids to enhance elderly and impaired people's autonomy in everyday life; autonomous wheelchairs, mobility aid lifters, feeding aids and rehabilitation robots allowing tetraplegics to perform manual labor tasks are examples of such service devices. In perspective, other than an all-purpose robot waiter, assistance, and healthcare systems integrating robotic and telematic modules will be developed for home service management (domotics).

Several robotic systems are employed for medical applications. Surgery assistance systems exploit a robot's high accuracy to position a tool, i.e., for hip prosthesis implant. Yet, in minimally-invasive surgery, i.e., cardiac surgery, the surgeon operates while seated comfortably at a console viewing a 3D image of the surgical field, and operating the surgical instruments remotely by means of a haptic interface.

Further, in diagnostic and endoscopic surgery systems, small teleoperated robots travel through the cavities of human body, i.e., in the gastrointestinal system, bringing live images or intervening in situ for biopsy, dispensing drugs or removing neoplasms.

Finally, in motor rehabilitation systems, a hemiplegic patient wears an exoskeleton, which actively interacts, sustains and corrects the movements according to the physiotherapist's programmed plan.

4. *Human-robot interaction*

We realise that a new gadget has become a daily fixture when no one is amazed by its presence in our environments. When any new invention first entered our lives, all new technologies provoked strong reactions such as terror, admiration, idolatry or aversion. Locomotives, cars, personal computers and mobile phones had to wait many years before they became accepted in our lives. It would seem that the next technology which is the candidate to become pervasive in our daily lives is robotics. Paradoxically, it has been held back by fears of excessive artificial intelligence and science fiction. Many domestic robots are actually on the point of being ready for mass use, and numerous research centres suggest that mobile robot manipulators will enter our homes and offices

very shortly. However, at the moment, there are only a few systems on the market.

The extension of robotic applications from the manufacturing industry to a daily life context is increasing as a result of the progressive lengthening of life expectancy in the more industrialised nations, as well as simplifying some day-to-day tasks. In the western countries, robots fit into the slot of improving our quality of life, entrusting hard or repetitive jobs to them. In Japan, instead, many robots are being developed as play-friends for children or carers for the old, as, for example, the humanoid robots and zoomorphical robots. The Shinto and Buddhist religions believe that even machines have souls, and this belief may have played a significant role in their positive acceptance by Japanese people as personal assistants.

On the international committees of the recent discipline of roboethics, this aspect is discussed with great attention, and the enormous US investment into research into robotics for military application is observed with concern. The robot soldier removes the final deterrent of war: the loss of troops at the front line. However, the autonomy of robotic soldiers in the critical functions of military objective selection and targeting has raised substantive concerns about the respect of International Humanitarian Law (IHL) and the possibility of identifying responsibilities for its violation[4] (Amoroso and Tamburrini. For a robot which has to interact closely with humans, however, there is a fundamentally valid condition in the use of industrial robots, that is the segregation between workers and production lines using robots, separated by barriers: now there is a need for robots capable of interaction with humans.

At the moment, the interaction with robots is really very dangerous, and there are no standard criteria of safety, nor is

[4] Cfr. D. Amoroso, G. Tamburrini, *The Ethical and Legal case against autonomy in Weapons Systems*, in «Global Jurist», 17, 3, 2017, pp. 1-20.

research into natural voice-operated interaction at such a point that a robot can be stopped in a case of emergency. The two key words are therefore safety and reliability. Numerous solutions to guarantee an increased concentration on these needs have been proposed over the last few years, but we have observed a lack of regulation, and the problem of combining safety with the traditional criteria of optimum functioning of a robotic system (speed and accuracy) is still an unsolved challenge. A robot is capable of using immense force to complete heavy tasks. If it is necessary to create great power to meet these human physical limitations, then safety is put at risk by the forces involved.

Up until today, a sort of Cartesian dualism (and corresponding division of scientific labour) has stated the dichotomy between mind and body of robots, entrusting the study of the former to neuroscientists and computer scientists, and the study of the mechanical structure and its control to the electronics, mechanical and cybernetics engineers. Now, in the present applications of robotics, we can see how the physical perspective has become a priority and thus the design of robot controllers cannot be independent from its physical structure.

Safety and reliability, therefore, must be placed in relation to the single components of the building of a robot, from the mechanisms to the motors, from the sensors to the control systems, understanding how malfunctioning and errors can be transformed into unexpected movements and collisions. The automobile industry is the first sector where studies are being made into quantitative measurements to evaluate the consequences of eventual accidents on the users of a mechanical system in movement (the passengers of a car). Some of the results can be used to define the thresholds, in terms of impact forces, beyond which the collisions can be considered fatal for an operator interacting with a robot. Levels of seriousness of the impact on a skull, for example, can be used to limit the velocity and acceleration of a mobile and

manipulative robot, but clearly the existing criteria have to be adapted.

In the near future, quantitative measurements should be introduced also in relation to the safety of a closer type of interaction with robotic systems, such as artificial intelligence and the responsibility of designers. Every technology must come to terms with a minimisation, as far as is possible, of situations which can lead to possible risk: many researchers in the Italian and European robotics community are dedicating themselves with great interest to the study of the problem of safety of robots in the home. Limited to an approach in which there are no invasive interfaces, and the interaction is external, they will have to draw up laws for the control of robots in such a way that they will not harm the users during normal functioning.

The fans of science fiction will remember Asimov's three laws for robotics, according to which a robot was obliged to functioning in such a manner to not harm a human (first law), obey human commands (with respect to the first law) and preserve itself (with respect to the previous two laws). It is clear that we cannot delegate everything to a central intelligence of robots: Asimov's laws are science fiction because it is not possible to understand the will of a robot, nor is it possible to avoid misunderstandings in the reasoning of an intelligent system: a robot may be quite unaware of the harm which it is causing. It is clear that the physical dimensions become also more important than the cognitive aspects (above all in cases of autonomous behaviour of robots), because unexpected movements of people can have tragic consequences. In any case, cognitive aspects are fundamental to give robots invasive interfaces and systems of sensorial fusion, which make them more aware and adaptable to interaction with people.

We have to act on practically every component of a robot: we need innovative materials for the mechanical structure, just as we need passive protection and control instruments against collisions and

to manage the successive phases and eventual impact. The design plans must include the possibility of dealing with errors in the various components in order to make them non-catastrophic, and sensor systems must offer a faithful image of position, direction and eventual expression of the voice-activated commands of the people present in the work environment. Finally, motors and activating systems of hand movements must not harm a user and aid movements and intentions.

5. *Robotics for Industry 4.0*

No doubt, in the last few years artificial intelligence (AI) has become the keyword which defines the future and everything that it holds. Not only has AI taken over traditional methods of computing, but it has also changed the way industries perform. From modernising healthcare and finance streams to research and manufacturing, everything has changed in the blink of an eye. AI has had a positive impact on the way the Information and Communication Technology (ICT) sector has developed. Looking ahead, however, the further growth of the ICT sector might experience a sort of saturation. With the advent of Cyber-Physical Systems, as in the Industry 4.0 programme in Europe, new enabling technologies such as 3D printing and robotics have opened a new prospect for a gradual and radical transformation from ICT to InterAction Technology (IAT), where the 'A' is intentionally capitalised to emphasise the importance of the physical action.

With the massive and pervasive diffusion of robotics technology in our society, we are heading towards a new type of AI, which we call Physical AI at the intersection of Robotics with AI, that is the science of robots and intelligent machines performing a physical action to help humans in their jobs of daily lives. The robot has *de facto* transformed into a cobot. A cobot is a robot actively cooperating with humans. The distinctive features of a cobot are:

- It can be used safely in a space shared with humans
- It has intuitive programming and communication interfaces
- Often it has particular physical characteristics, it is equipped with exteroceptive sensor and an advanced control system

The physical characteristics in the design of a cobot are:

- lightweight and transportable
- redundant
- double arm
- free of edges
- covered with padding

while these are its typical sensors:

- joint torque sensors
- wrist force/torque sensors
- 3D vision
- sensitive "skin"

and its control modes:

- impedance control
- collision detection
- human–robot interaction

As far as programming a cobot, further to traditional on-line lead-through programming (with tech pendant) and off-line programming, one has intuitive programming modes:

- on-line walk-through (manual guidance)
- programming by demonstration
- virtual and augmented reality
- multimodal communication (gestures, voice, touch)

Within the Industry 4.0 framework, new designs are aimed at making robots and cobots customisable machines which could be intuitively operated even by unexperienced users according to a plug-and-play paradigm. Physical assistance to disabled or elderly people;

reduction of risks and fatigue at work; improvement of production processes of material goods and their sustainability; safety, efficiency and reduction of environmental impact in transportation of people and goods; progress of diagnostic and surgical techniques are all examples of scenarios where IAT is indispensable.

6. *Roboethics and human–robot interaction in the workplace*

During the second half of the 20th century, robotics technologies and systems greatly contributed to reshape industrial production. Present and foreseeable advances in robotics research promise to have an even more profound impact on human working activities, by reshaping highly specialised working activities – in medical, personal care and other professional domains – and by paving the way to the new forms of human–machine cooperation and interaction that are required by Industry 4.0 innovation plans.

The continuing impact of robotics on working conditions and activities raises a variety of significant ethical issues that must be properly analyzed and addressed. These issues arise against the background of a variety of normative ethical principles concerning human work and what one ought to do in that application domain for robotics. Providing a complete list of such principles is a daunting and possibly unachievable task, considering the plurality of theories in normative ethics and their historical developments. However, one can hardly doubt that the following ethical principles play a crucial role in the context of robotics applications in the workplace:

1. Human responsibility principle: prospective and retrospective responsibilities for the activities of robotic systems, including the responsibility to protect the human body in physical human–robot interactions, should be fairly distributed among human agents.

2. Distributive justice principle: the wealth produced by means of robotic systems should be fairly distributed.
3. Dignity respect principle: the dignity of human work as such, and human dignity in human–robot working interaction should be respected.

These and other principles for moral judgment and action do not come with a recipe that one applies mechanically to derive ready-made solutions to each specific moral problem. Rather, one must think through each moral problem under scrutiny, with the aim of evaluating the relevance of these normative principles, interpreting them in context and figuring out their situational implications.
From a methodological viewpoint, ethical problem-solving in roboethics is profitably viewed as a reflective activity on specific moral issues guided by these and other general normative principles, and involving two major stages: (i) identification and analysis of ethical issues concerning some specific class of robotic technologies and systems in the light of general ethical principles; (ii) development, based on this analytical work, of ethical policies for the design and use of those technologies and systems.
In the following section the relevance and applicability of the first ethical principle listed above is dealt with in the context of surgical robotics, especially in the light of technological advances towards increasingly autonomous surgical robots that are contributing to reshape further the highly specialised working activity of human surgeons. In the final section the relevance and applicability of Principles 2 and 3 is framed within the context of discussions about technological unemployment, if any, caused by robotisation of work tasks, and of Principle 3 –but only more briefly so– in connection with working conditions in human–robot cooperative teams.

7. Increasing autonomy of surgical robots and human responsibility

The responsibility principle listed above requires one to distribute fairly among human agents the prospective and retrospective responsibilities for the actions of robotic systems, including the responsibility to protect the human body in physical human–robot interactions. The interpretation of this principle raises special ethical issues in the context of the increasing autonomy of medical robots, where physicians are no longer in control of each and every aspect of medical procedures on the human body. A schematic hierarchy of six autonomy levels for medical robots was introduced by Yang and co-authors[5]. Starting from medical robots having no autonomy (L0 autonomy), at the next levels of this hierarchy one finds robotic assistants constraining or correcting human action (L1), robotic systems carrying out tasks that humans designate and supervise (L2), and robotic systems additionally generating task execution strategies under human supervision (L3). The proposed hierarchy is rounded out by technologically more distant robotic systems performing an entire medical procedure with or without human supervision (L4 and L5 respectively).

Contextually to the introduction of this hierarchy, Yang and co-authors advanced the requirement that treating physicians should be «still in control to a significant extent». A robust motivation for this requirement is found in the human responsibility principle stated above, which additionally enables one to clarify more precisely what it means to be «still in control to a significant extent». Indeed, the principle entails that human control over increasingly autonomous medical robots should enable one to prevent or reduce damages that medical robots may bring about (prospective human responsibilities). And the principle additionally entails that human control should be designed so as

[5] G. Z. Yang *et al.*, *Medical robotics-regulatory, ethical, and legal considerations for increasing levels of autonomy*, in «Science robotics», 2, 4, 2017.

to avoid responsibility gaps when these damaging events do occur and to enable the distribution of moral and legal responsibilities among involved human actors. When these conditions are satisfied, one may justifiably assert that meaningful human control (MHC) over robotic autonomy is put in place[6].

On the basis of these background observations, a specialisation of the above autonomy levels hierarchy to surgical robots is addressed[7], along with the related problem of establishing MHC over robots at each level in this hierarchy.

In the medical domain of Robot-Assisted Surgery (RAS), L0 autonomy systems are used as slave devices for scaling motion, attenuating tremor and enhancing the precision of surgical gestures. Indeed, the da Vinci robotic system for laparoscopic surgery is typically configured as a teleoperated system with L0 autonomy, where surgeons exercise direct control over the entire surgical procedure, including data analysis, preoperative and intraoperative planning, decisions and actual execution. Clearly, the MHC requirement flowing from the human responsibility principle above is unproblematically satisfied when these settings are in place.

More subtle MHC issues arise at L1-L3[8]. Various surgical robots deployed in operating rooms are already granted L1 autonomy. A significant case in point are robotic systems assisting surgeons to move the manipulator along desired workspace paths or preventing robotic manipulators from entering selected workspace regions. Robotic systems identifying and applying these active constraints (aka as Virtual Fixtures) are more than slave devices,

[6] D. Amoroso, G. Tamburrini, *I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*, in «Biolow Journal», 1, 2019, pp. 33-51.

[7] M. Yip, N. Das, *Robot autonomy for surgery*, in R. Patel (ed.), *The Encyclopedia of Medical Robotics*, World Scientific, Singapore 2018, pp. 281-313.

[8] F. Ficuciello, G. Tamburrini, A. Arezzo, L. Villani, B. Siciliano, *Autonomy in surgical robots and its meaningful human control*, in «Paladyn Journal of Behavioral Robotics», 10, 2019, pp. 30-43.

as they on occasion correct the surgeon's intended motions. Therefore, to exert MHC at this autonomy level, one must have the option to override robotic corrections, by means of second-level human control privileges enabling the surgeon to prevail on first-level robotic corrections.

At L2, humans select a task for surgical robots to perform. The surgeon's supervising role consists in hands-free monitoring and possible overriding of robotic execution. Thus, the robotic system is under the surgeon's discrete (rather that continuous) control. The ROBODOC system for orthopaedical surgery is a relatively early example of a system deployed in operating rooms and endowed with L2 autonomy, insofar as it carries out bone milling preoperative plans under human supervision. A more recent research prototype endowed with L2 autonomy is the experimental Smart Tissue Autonomous Robot (STAR) platform[9] which carries out intestinal suturing (anastomosis) on pig tissue. In experimental tests on this animal model, STAR was found to outperform expert human surgeons in manual laparoscopic surgery conditions on account of various clinically used suturing metrics.

The ROBODOC and STAR surgical systems are presently characterised by different Technology Readiness Levels (TRLs). The former system is used for clinical standard procedures, while the latter is still at the research level. This disparity crucially depends on the nature of their respective operational environments and predictability properties. ROBODOC's surgical sites are rigid anatomic structures, whereas STAR operates on deformable soft tissues. The structured environments where ROBODOC operates allow for safe autonomous task execution due to the possibility of making accurate measurements and scene changes predictions. In contrast with this, the soft and deformable surgical sites where STAR operates raise more severe challenges for the accurate

[9] A. Shademan, R. S. Decker, J. D. Opfermann, S. Leonard, A. Krieger, P. C. W., *Supervised autonomous robotic soft tissue surgery*, in «Science translational medicine», 8, 2016, pp. 337-364.

detection and tracking of both surgical tools and anatomical parts. These differences in the ROBODOC and STAR operational environments suggest that the human perceptual and cognitive vigilance must be suitably modulated to achieve MHC of individual surgical robots that one brings together under the broad category of L2 autonomous robots. Discrete perceptual sampling and cognitive evaluation of robotic task execution are arguably more demanding in the case of STAR-like systems, in view of scene changes due to physiological blood flow and respiration, and the corresponding need to assess the robot's adaptive response. Accordingly, one size of discontinuous MHC control does not fit all L2 autonomous surgical robots.

L3 autonomous surgical robots generate task strategies under human supervision, and conditionally rely on humans to select from various generated strategies or to approve an autonomously selected strategy. To a limited extent, STAR achieves this level of conditional autonomy as far as anastomosis strategies generation is concerned, along with systems dynamically identifying virtual fixtures and generating optimal control parameters or trajectories.

MHC for L3 autonomy distinctively requires surgeons to decide competently whether to approve one of the robot generated strategies. This decision presupposes that surgeons understand the rationale for proposed strategies, are in the position to compare their respective merits, and to make up their mind in due time about which strategy to prefer over alternatives. Depending on the complexity of proposed strategies and surgical sites, MHC may incrementally raise human interpretability and decision-making challenges about robot generated strategies. Similar issues may emerge in connection with strategies that surgical robots may learn to propose on the basis of machine learning methods, in view

of interpretability problems affecting learning systems[10]. Today, the learning of surgical strategies is bound to be based on data sets formed by humanly generated strategies. In a more distant future, interpretability and explanation issues arising in the context of MHC for level 3 robotic autonomy may become increasingly acute if datasets for learning how to generate intervention strategies progressively shift from data concerning human-generated strategies to robot-generated strategies and corresponding clinical outcomes.

Schematically, to identify proper MHC policies for surgical robot autonomies one has to consider the functionalities that are appealed to define hierarchies of increasingly autonomous surgical robots (the what of autonomy), the bodily environments in which these robots operate (the where of autonomy), and the system capabilities that are deployed, e. g. learning, to undertake given autonomous actions (the how of autonomy). From an ethical standpoint, the identification and application of MHC policies on increasingly autonomous surgical robots is motivated by the bioethical principles of beneficence and non-maleficence[11] in general, and by the prospective deontological responsibilities of surgeons that these principles entail.

A thorough analysis of prospective responsibilities induced by the MHC requirement is needed to address the problem of developing suitable training programs for surgeons in RAS. In particular, the non-maleficence bioethical principle requires proper training to provide conceptual tools countervailing positive machine biases, which may wrongly induce human surgeons to trust more what the robot does or proposes to do rather than their own contrasting judgment. A thorough analysis of MHC-related duties plays an equally significant role in evaluating what are the surgeon's

[10] G. Montavon, S. Wojciech, K. R. Müller, *Methods for interpreting and understanding deep natural networks*, in «Digital Signal Processing», 73, 2018, pp. 1-15.

[11] T. L. Beauchamp, J. F. Childress, *Principles of Medical Ethics*, 7th edition, Oxford University Press, Oxford 2013.

retrospective responsibilities, if any, when something goes wrong. Indeed, a surgeon might be held responsible for damages caused by an autonomously performing robot if she failed to exert MHC properly and the harm in question might have been averted had she carefully complied with her MHC duties. By the same token, retrospective responsibility allegations against surgeons for damages caused by an autonomously performing robot might be rebutted and possibly diverted towards other human agents by showing that the specified MHC duties were judiciously complied with.

8. *Roboethics and technological unemployment*

The distributive justice principle is considered in the context of possible (but as yet unobserved) long-term technological unemployment effects flowing from the robotisation of many working tasks and activities, ranging from routine manual tasks of assembly lines to highly specialised tasks involved in surgical interventions. Industrial robots are the largest commercial application of robotics in industrial manufacture. Robots are taking on working roles in agriculture and forestry, construction, mining, exploration of hazardous environments, rescue operations and disaster response[12]. Moreover, it was pointed out above that increasingly autonomous robots are bringing about major changes in transportation and logistics, healthcare and personal assistance, defence, surveillance and security. And more distant visions mentioned above suggest that robots will additionally pervade domestic life, adding to the initial functions of home cleaners the more challenging activities of dexterous assistants, helpers and tutors.

In the light of these advancements and forecasts, robotics is expected to create new markets while displacing established markets and firms, thereby playing the role of a major disruptive

[12] Cfr. B. Siciliano, O. Khatib, *Springer Handbook of Robotics*, cit., Part F.

technology in the 21st century. In this economic process of creative destruction, robotic innovation is expected to affect the nature of many jobs, to displace various human occupations, and to generate new job opportunities. A question naturally arising in this framework is whether robotic innovation will cause widespread and lasting unemployment. Will there be enough new jobs to replace jobs that disappear on account of robotic automation? Similar questions about technological unemployment emerged throughout the history of technological innovation: from mechanised looms introduced in textile manufacturing at the end of the 18th century to the automation of car manufacturing, starting from Ford's moving assembly line in the early 20th century and leading in the early 21st century to the highly automatised Daimler Factory 56 in Sindelfingen (Germany).

According to a traditional macroeconomics model, one should worry about the social implications of technological unemployment for short periods only in the wake of major episodes of technological innovation. This model predicts that increased productivity induced by automation will reduce the price of goods; that wages will accrue greater purchasing power on this account, thereby stimulating the demand for new goods and services; and that new economic activities will be created to satisfy this demand. Many past episodes of technological innovation fit into this model of displaced jobs that are eventually outnumbered by newly created jobs and increased wealth benefiting large social groups. However, the future predictions of this "virtuous circle" model about the positive effects of robotisation and computerisation in the XXI century were questioned in the wake of academic studies about the sheer quantity and variety of manual and intellectual tasks that

are likely to be automated, and specifically so on account of imminent advances in both AI and robotics[13].

Less alarming outlooks were made in later economic studies. In the more recent OECD report entitled the future of work[14], for example, it is stated that automation may cause about 14% of existing jobs to disappear in the course of the next 15-20 years; more than 30% of the other jobs will undergo a radical transformation. At the same time, new temporary and less well-paid jobs will emerge, for a variety of reasons which do not necessarily have to do with robotisation or computerisation of working tasks. The OECD report is careful to emphasise that the benefits that may flow on the basis of the "virtuous circle" model of automation may occur on an extended temporal scale, which is inadequate to respond to the more pressing needs of those who become unemployed for reasons which may include globalisation, demographic changes, but also short-term effects of automation: «The future of work offers unparalleled opportunities, but also significant challenges. Globalisation, technological progress and demographic change are having a profound impact on society and labour markets. It is crucial that policies help workers and society at large to manage the transition with the least possible disruption, while maximising the potential benefits»[15].

It is not a proper concern for roboethics to adjudicate these macro-economic predictions and disputes. However, roboethics is definitely concerned with a related normative question: Is there a social duty to act and countervail lasting job losses in case a pessimistic outlook about the implications of robotic innovation for human labour comes true?

[13] C. B. Frey, M. A. Osborne, *The future of employment: how susceptible are jobs to computerization*?, in «Technological Forecasting and Social Change», 114, 2017, pp. 254-280.
[14] OECD, *The future of work. Employment Outlook 2019*, Organisation for Economic Cooperation and Development, http://www.oecd.org/employment/outlook/.
[15] *Ibid.*

To address this normative question, one may draw on theories of social justice and related conceptions of equality, desert, and entitlement. To illustrate, consider the implications of Rawls's influential theory of justice as fairness in a scenario of persistent technological unemployment hypothetically due to robotics and related AI innovations. According to this theory, human beings are entitled to certain primary goods in order to develop their rational plans of life. These primary goods include self-respect, in addition to «rights and liberties, powers and opportunities, income and wealth»[16]. Thus, justice as fairness urges one to contrast the loss of earned income that one needs to develop rational plans of life, or to compensate for this loss in order to ensure the provision of primary goods by other means. Likewise, earned income is an instrument for developing human capabilities and achieving satisfactory human living according to so-called capability approaches to justice[17]. Hence, capability approaches to justice require one to neutralise impediments to the flourishing of individual human capabilities possibly deriving from technological unemployment.

Distributive principles –and their moral grounding in duties to supply primary goods, foster human capabilities or enhance welfare– jar with the economic freedom of persons that some liberal thinkers prioritise. According to von Hajek, the very idea of distributive justice is based on a categorical mistake, because neither society not its institutions are moral agents which one may call just or unjust[18]. A forceful rejoinder to von Hajek's objection is based on the observation that at least in democratic societies individual moral agents can make coalitions and support policies that are coherent with their shared moral conceptions. As a champion of liberalism, Hajek additionally claimed that public

[16] J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge 1971, p. 62.
[17] A. Sen, *The Idea of Justice*, Allen Lane, London 2009, ch. 12.
[18] F. A. von Hayek, *The atavism of social justice*, in *New Essays in Philosophy, Politics and Economics*, Routledge and Keagan Paul, London 1978.

redistribution of wealth limits individual freedom and creates inefficient distortions of the market economy, whose unperturbed developments is expected to benefit everybody in the long run. In particular, redistribution interventions may stifle technological innovation and the social benefits that come with it. Familiar economic objections to this ideal view of market self-regulation are based on recurring market failures in the 20th and 21st centuries.

As in many other cases of interest to roboethics, these sketchy remarks on distributive justice debates show that there is no guarantee to converge on a consistent set of moral directives about the distribution of wealth created by means of robotic automation and hypothetical scenarios of persistent technological unemployment. Accordingly, public discussion and deliberation is needed here too, to achieve a proper balance between personal economic freedom, the social benefits flowing from social innovation and distributive justice concerns about short-term (or even long-term) technological unemployment.

The above discussion bears on the ethical issue of dignity of human work as such, and thus on a contextualisation to robotics in the workplace of the first part of the dignity respect principle. The second part of this principle has to do with the respect of human dignity in human–robot working interaction. Thus, in addition to issues concerning the dignity of human work as a source of earned income, roboethics must be concerned with issues of dignity at work. The latter depends on workers' autonomy and self-mastery in working activities, on self-esteem flowing from their contributions to the value chain of their organisation, on workplace interactions promoting trust, recognising competence, and offering the opportunity of being respectfully listened to[19] (Sawyer 2007). Issues of dignity at work that are specific to

[19] A. Sawyer, *Dignity at work: broadening the agenda*, in «Organization», 14, 2007, pp. 565-581.

robotics must be addressed already at the design stage of mixed human–robot cooperative work and teaming, by proper allocation of decision-making authority and distribution of tasks.

Acknowledgement
The contributions by Daniela Passariello to the concepts discussed in this article are gratefully acknowledged.

BRUNO SICILIANO is professor of Control and Robotics, responsible of PRISMA Lab and director of ICAROS Center at University of Naples Federico II. Honorary professor at Óbuda University, past-president of IEEE Robotics and Automation Society, he has been funded 20 European projects in the last 12 years, and has received several international awards and prizes

bruno.siciliano@unina.it

GUGLIELMO TAMBURRINI (PhD 1987, Columbia University) is professor of Philosophy of Science and Technology at University of Naples Federico II. Coordinator of the first European project on ethics of robotics, recipient of Gulio Preti International Prize, he is member of the International Committee for Robot Arms Control (ICRAC)

guglielmo.tamburrini@unina.it