On Ethical Questions in Robotics and Artificial Intelligence

Raja Chatila
Institut des Systèmes Intelligents et de Robotique (ISIR)
UPMC - Paris, France
What is a Robot?

- A physical machine endowed with three basic capacities:
  - Data acquisition through sensors and interpretation of data to extract information and knowledge.
  - Decision-making: to determine and plan a course of action, to achieve an objective or to react to events.
  - Action and motion in the real world through actuators.

- Two additional capacities
  - Communication with operators, users, or other machines.
  - Learning to improve world representations or action performance from experience.
Remarks

- These capacities may be developed up to various degrees of complexity.

- They may be integrated in a single “body” or be distributed (i.e., perception or decision making may be separate and remote from the operational functions).

- We focus on single robots, but one could also consider interacting multi-robot systems.
Autonomy

- Ability of a system to decide without the assistance of another agent.
- Related to the aforementioned capacities and their integration.
- Operational autonomy:
  - Related to data processing to build basic representations such as terrain models, and to control motion/action, e.g., to avoid obstacles.
  - Widely present in today’s deployed system
- Decisional autonomy: Related to reasoning on perception and action, to assess complex situations and take non trivial decisions.
State of Robotics Research and Applications

- Robotics research has reached a certain level of maturity to achieve operational functions in perception, motion planning, control, human-robot interaction, ...

- Autonomous robots are a paradigm of AI, with capacities of decision making and action planning.

- Autonomous operation (e.g., navigation) is yet possible in only some specific situations (e.g., limited terrain).
Robot Applications: manufacturing, hazardous environments, transport, defence, entertainment, agriculture, construction, health

- To replace humans

- To assist and help humans

- To serve humans

- To rehabilitate/augment humans
Ethical Awareness

- Questions on ethical and legal issues in the use of robots have emerged and are becoming important to scientists and to the public.
- Autonomous robots raise additional questions mainly in military applications (Autonomous Lethal Weapons).

- The general public is often not aware of the actual state of the art in R&AI, which is rather limited, but expresses concern.
Scientists have started to reflect on the question of the ethical implications of robotic technology and of autonomous robots more than ten years ago (first workshop: 2004).

Motivated by reflections on moral and legal responsibilities of scientists

Ethical, Legal and Societal issues (ELS) addressed in several robotics projects in the EU (FP6, FP7)

International conferences, workshops and working groups.

Initiatives toward forming ethics committees on robotics (similar to Biology).
A few ELS Issues

- Jobs,
- Accountability and responsibility,
- Autonomous robot decisions,
- Privacy, intimacy, intrusion, surveillance,
- Human dignity,
- Dependence, isolation of people,
- Cognitive and affective bonds,
- Bio-mimicry,
- Human identity,
- Human augmentation,
- Human ethics in using robots,
- Status of the robot in the human society
CERNA

- French advisory commission for ethics of ICT research (Commission de Reflexion sur l'Ethique de la Recherche en Sciences et Technologies du Numérique).
- Formed in October 2012.
- First case study: Robotics.
- Issued recommendations on the ethics of research in Robotics in November 2014.

Approach of the CERNA

- Impossible to predict all applications and their consequences
- Adopt a perspective on the ethics of research
- Case domain examples raising ethical issues:
  - Defence and security
  - Rehabilitation and human augmentation
  - Assistant robots for vulnerable people
  - Robot companions and robots for the general public
- 4 Robot capacities raising ethical issues, and recommendations:
  - Autonomy
  - Human augmentation
  - Emotions and bonding
  - Bio-mimicry
Robotic Applications Raising Ethical Issues - Defence and security

- Drones, swarms, autonomous robots, UGVs, in defence and security

Predator

Crusher CMU

IRobot Packbot

Drones for security monitoring
Robotic Applications Raising Ethical Issues - Rehabilitation and human augmentation

- Robotic devices for rehabilitation - Human augmentation

Ecso Bionics

U. Saragozza

RIC Institute

RB3D
Robotic Applications Raising Ethical Issues - Assistive robotics for vulnerable people

- Assistive robots for vulnerable and fragile persons (elderly, children, handicapped people).

- Paro
- Telenoid - Osaka U.
- S. Korea
- Aldebaran Romeo
Robotic Applications Raising Ethical Issues - Robot companions and robots for the general public

- Robot companions, personal assistant robots, sexual robots.

Aldebaran Nao

The Big Bang Theory
Robotic Applications Raising Ethical Issues - Robot companions and robots for the general public

- Autonomous cars (Legal and social issues, ethical dilemmas)
Focus 1: Robot Autonomy and Ethics

- Issues
  - Autonomous decision-making vs. Autonomous operation
  - Robot situation awareness and interpretation
  - Generally: Robot+Operator in a shared autonomy
  - Human awareness of robot state; surprises.
  - Human over-confidence in robots
  - Moral buffer
  - Robot responsibility vs. human responsibility.
Robot Autonomy in Operation

- All autonomous systems are supervised by a human operator at some level. All autonomous systems are joint human-machine cognitive systems.

- Autonomy is not an intrinsic property of an unmanned vehicle in isolation: design and operation of autonomous systems need to be considered in terms of human-system collaboration.

- System autonomy: a continuum from complete human control on all decisions to situations where many functions are delegated to the robot with only high level supervision and/or oversight from its operator.

- Some functions may require a human in the loop whereas others can be delegated at the same time.
Autonomy and Authority Sharing

- **Machine**
  - Limited capabilities: does what it is programmed to do
  - Uncertainties in perception and interpretation (e.g., combatant/non-combatant); in action results (e.g., slippage)
  - Hazards: failures, environment changes

- **Operator**
  - Busy, fallible, stressed (e.g., attentional tunnelling)
  - Moral buffer - human moral distance wrt. machine actions
  - Procedures not strictly followed

- **Interaction**
  - Communication perturbations
  - Automation bias: over-confidence in machines
  - Automation surprises: discrepancy between situation awareness and actual system state
Recommendations for Research on Autonomy - 1

{AUT-1} Control
Researchers should investigate the capacity of the operator or the user to take over control from the robot and that of the machine to take over control from the human, including the circumstances when such takeover is allowed or mandatory. Researchers should investigate whether the human is to be allowed to disengage autonomous robotic functions.

{AUT-2} Decisions made without operator’s awareness
Researchers must ensure that robotic decisions are not made without operator’s knowledge, in order to avoid gaps in the operator's situational awareness. It is imperative to ensure that the operator will never believe the robot to be in a certain state while in fact it is in a different state.

{AUT-3} Effects on operator’s behaviour
Researchers should be aware of the trust bias, i.e., operator’s tendency to exhibit excessive confidence in robotic decision-making procedure, and of the moral buffer, i.e., operator’s tendency to morally disengage from robotic actions or behaviour.

{AUT-4} Programming limits
Researchers should evaluate interpretative and decision-making software and understand and be able to explain its limits. In particular, whenever a robot is to be endowed with moral behaviour, the limits of programming should be carefully assessed.
Recommendations for Research on Autonomy - 2

{AUT-5} Situational awareness
With regard to interpretative robotic software, researchers should evaluate the extent to which it can correctly characterize a situation and distinguish between apparently similar situations, in particular if this characterization is the only basis of the ensuing decision or action. It is also necessary to evaluate the methods of accounting for uncertainties.

{AUT-6} Predictability of a human-robot system
More generally, researchers should analyse the predictability of a human-robot system by considering uncertainty in interpretation and action, possible robotic or human failures, and the entire set of states that can be reached by the system.

{AUT-7} Traceability and accounting
Researchers should develop tracing tools at the design stage of a robot. These tools will facilitate accounting and explanation of robotic behaviour, even if limited, at the various levels intended for experts, operators and users.
Focus 2: Human Augmentation and Rehabilitation

- Issues
  - Rehabilitation vs augmentation?
  - Dignity (gain or loss)
  - Privacy (data).
  - Status of the augmented human
Recommendations on Human Augmentation and Rehabilitation

{RAH-1} Medical ethics
Researchers in reparative or assistive robotics should, in coordination with healthcare professionals and patients, apply the principles of medical ethics in order to make informed choices between the requirements of care efficacy and safety, patient independence and integrity, and privacy protection. These questions should not only be considered from the legal standpoint; ethical thinking and deliberation help to make individual adjustments on a case-by-case basis rather than apply a general rule. Researchers should solicit and follow opinions published by operational medical ethical committees in view of establishing a connection between emerging robotic technology and positions expressed in such opinions.

{RAH-2} Individual independence and integrity
Researchers working on reparative robotic systems should seek to preserve independence of equipped individuals by maintaining them in position to control their actions as far as possible. Researchers should also seek to preserve the integrity of functions other than those being repaired.

{RAH-3} Reversibility
Researchers intentionally working on robotic devices for human enhancement must ensure that the resulting augmentation remain reversible. Devices should be removable without causing harm to the person and without loss of the initial functions.

{RAH-4} Societal effects of enhancement
Researchers should investigate societal effects of human enhancement induced by the devices they develop, including effects on the social behaviour of equipped individuals and, reciprocally, on the social behaviour of the unequipped persons.
Focus 3: Affectivity, Emotions, Bonding

- Issues
  - Bonding and Isolation of humans
  - Dependance for vulnerable persons
  - Understanding the status and capacities of the robot
Focus 4: Affectivity and Bio-mimicry and

- Bio-mimicry of aspect, behaviour, emotions
  - Expression of robot emotions
  - Human identity vs. androids, robots vs. living beings
  - Status of the robot in Human Society
Recommendations for Research (Affectivity and Bio-mimicry) - 1

{IVI-1} Utility and necessity in view of purpose
With regard to the useful and the necessary robotic functions, researchers should study relevance and necessity of provoking emotions and of biomimetic behaviour or appearance, in particular in the case of a strong visual or behavioural resemblance between a robot and a living being. When human voice or likeness are imitated, researchers should investigate the effects of such imitation, including those exceeding the sphere for which the robot is intentionally designed.

{IVI-2} Nature/artefact frontier
If a robotics project seeks quasi perfect resemblance between a robot and a living being, researchers should remain aware that biomimetic approach may blur the frontier between nature and artefact. In this case, researchers must consult an operational ethics committee of their institution.

{IVI-3} Study of the effects
In an affective robotics project, researchers should investigate all consequences that their work may have on the user socialisation capabilities.
{IVI-4} Child-robot interaction
In a robotics project that puts children in the presence of a robot, researchers should address the impact of child-robot interaction on the development of the child’s emotional capabilities, notably in early childhood.

{IVI-5} Evaluation
In a robotics project that may involve user affectivity, for example by seeking attachment to a robot, researchers should draw up design and evaluation protocols and join with potential users and stakeholders in the effort to make best informed scientific and technological choices.

{IVI-6} Communication
Researchers should exercise caution when they speak in public on robot emotions and on their resemblance with living beings. Researchers should remain aware that emotional expression by a robot is an illusion in the human sense and that the imitation of living beings may, intentionally or otherwise, facilitate the transfer of certain features of the living being to the artefact.
Some General Conclusions

- Define an ethics charter for Robotics research
- Reach an international consensus
- Correct balance between open research and ethical recommendations
- Set up operational ethics committees for research practice in Robotics.


(Will be shortly also available in English)