Ethicbots at the Euron Atelier on RoboEthics

Guglielmo Tamburrini
Università di Napoli Federico II

Genova, 2 march 2006

ETHICBOTS

- Emerging Technoethics of Human Interaction with Communication, Bionic, and Robotic Systems
- 6th Framework Programme
- Programme Science and Society
- Duration: November 1st, 2005 → October 31st, 2007

ETHICBOTS Consortium

1	University "Federico II", Naples, Department of Physical Sciences and Department of Computer and Systems Engineering	DSF	Italy
2	Fraunhofer Institute for Autonomous intelligent Systems, Sankt Augustin,	FhG/AIS	Germany
3	Scuola di Robotica, Genova,	SdR	Italy
4	Institute of Applied Philosophy, Faculty of Theology, Lugano,	IsFA	Switzerland
5	University of Reading, department of Cybernetics	UNIRDG	UK
6	Hochschule der Medien University of Applied Sciences, Stuttgart	HdM	Germany
7	LAAS-CNRS, Toulouse	CNRS	France
8	Scuola Superiore Sant'Anna, Pisa	SSSA	Italy
9	University of Pisa, Department of Philosophy	UNIPI	Italy
10	Middlesex University, Interaction Design Centre, School of Computing, London	UM	UK

ETHICBOTS

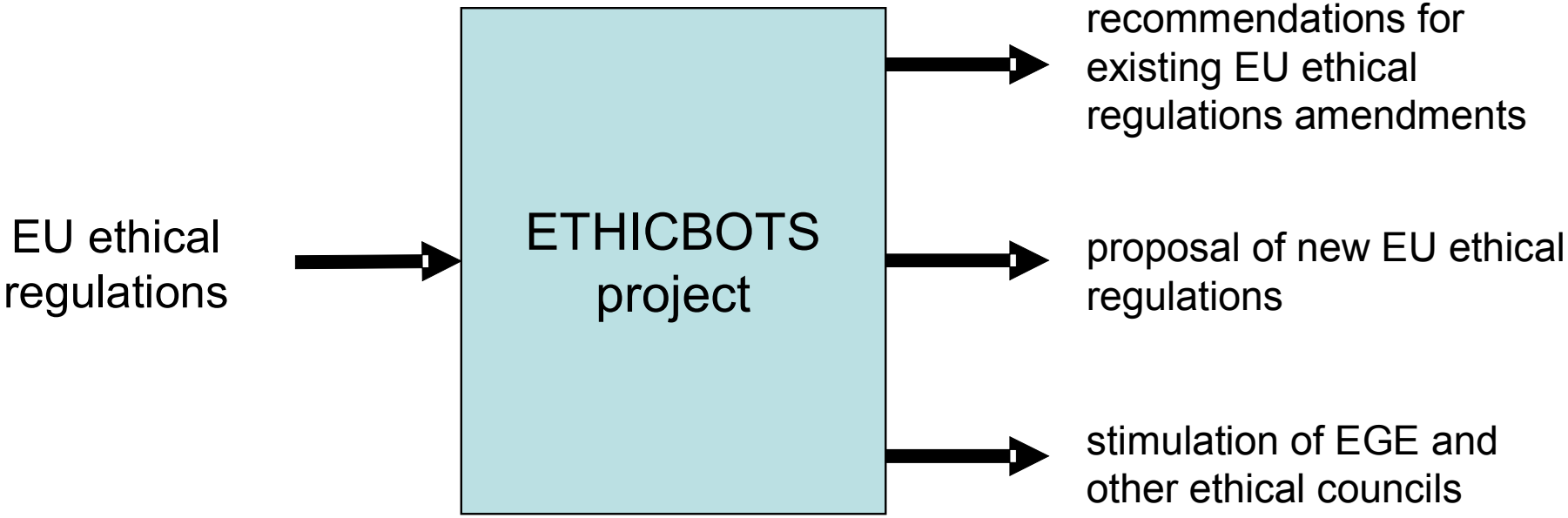
Strategic objectives

- ✓ **Raising awareness and deepening understanding of techno-ethical issues;**
- ✓ **Ethical monitoring of ICT, robotic, and bionic technologies for enhancing human mental and physical capacities;**
- ✓ **Fostering integration between Science and Society, by**
  - ✓ **promoting responsible research,**
  - ✓ **providing input to EU and national committees for ethical monitoring, warning, and opinion generation,**
  - ✓ **improving communication between scientists, citizens and special interest groups._**

Varieties of human-machine integration

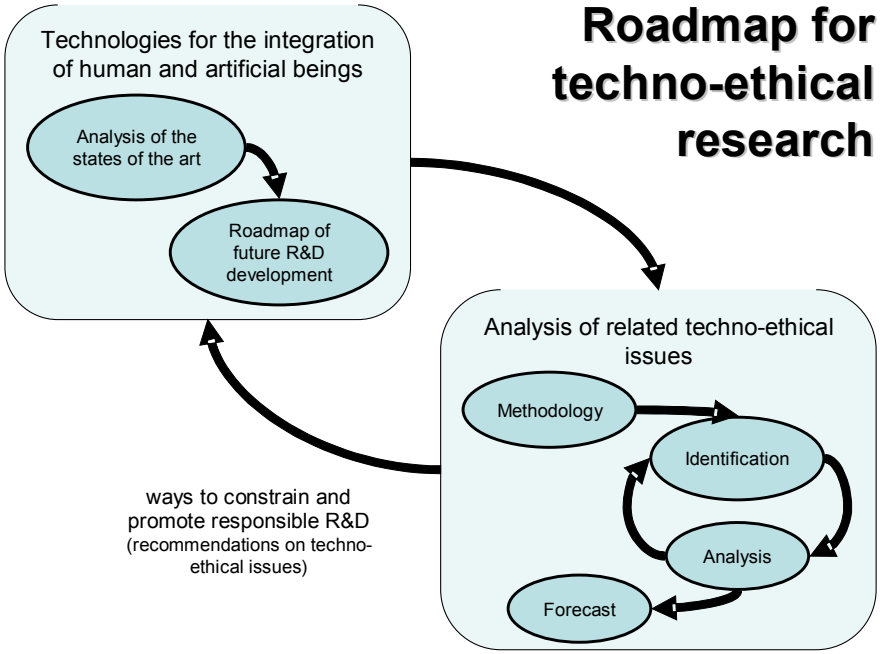
- *Human-softbot integration*: AI for information & communication.
  - distributed computing for human assistance
  - intelligent agents for information and communication management
  - edutainment AI
- *Human-robot, non-invasive integration*: Autonomous robotics.
  - personal and assistance robotics
  - edutainment robotics
  - warfare application of robotics
- *Physical, invasive integration*: Bionics.
  - prosthesis
  - enhancement of human sensorimotor capabilities
  - ICT implants

Contributions to standards



shifting attention from limitation to encouragement, guidance, and promotion of responsible scientific and technological research.

Roadmaps for techno-ethical research



The "Roadmap for Techno-Ethical Research" on the integration of human beings and artificial (hardware/software) entities will take the form of a comprehensive document (task 5.4; deliverable D10).

- It will be based on the charting of extant scientific and technological work in robotics, bionics, and Artificial Intelligence for information and communication systems, on projections about near future developments, on the triage of techno-ethical issues in view of European society objectives, needs, and problems

“Crude” ethical questions

- **who is responsible for actions carried out by human-robot hybrid teams?**
- **(how) is the ICT monitoring and use of personal data to be regulated?**
- **can bionic implants be used to enhance, rather than restore, physical and intellectual capabilities?**

From commonsense to critical ethical thinking

- Developments of robotics, bionics, and AI research undermine received views about human-machine interaction
- Interactions with robots raise issues about
  - protection of human freedom, dignity, and integrity,
  - promotion of human capacities and functions
  - justice, (fair access to resources),
  - human identity and autonomy
  - privacy, responsibility

Charter of the Fundamental Rights of the European Union

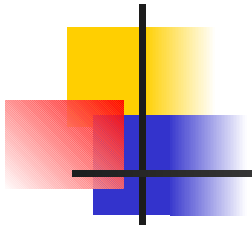
- Art. 1 (human dignity): human dignity is inviolable. It must be respected and protected.
- Art. 3: right to the (physical and mental) integrity of the person.
- Art. 6: right to liberty and security.
- Art. 8: right to protection of personal data.
- Art. 25: rights of the elderly to lead a life of dignity and independence and to participate in social and cultural life.
- Art. 26: Integration of persons with disabilities.

Conceptual analysis by experts

- Conceptual analysis on the basis of *specialized knowledge*
  - triaging techno-ethical issues,
  - deepening our understanding of the higher-ranked issues,
  - guiding ethically sustainable research,
  - identifying ethical motivations to promote new research perspectives.

Triaging: identifying potential impact categories

- Selecting Potential Impact Categories (PICs) as a basis for triaging emerging techno-ethical issues.
- Examples:
  - Imminence (science-fiction vs roboethics)
  - Novelty (lack of understanding and policies)
  - Social pervasiveness of technologies.



Dispelling misconceptions

"The machine will do exactly what we programmed it to do…"

- Do we fully understand the robots we make and theorize about?
- Can we fully predict and control robot behaviour?

Deepening our understanding
learning machines and responsibility

Designers, manufacturers, and operators cannot fully predict the behaviour of many learning machines based on

- symbolic learning
- neural network learning
- evolutionary algorithms

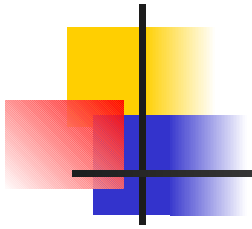
Ethical and legal responsibilities should be evaluated on the basis of this fact !

Moral and legal responsibility

- For a person to be held responsible, she must have control over her behaviour and the resulting consequences (Fisher and Ravizza, *Responsibility and Control*, Cambridge UP 1998, p. 13)
- But
  - one can hold a person responsible for her pet's behaviour, for exercising dangerous activities, for the ensuing damages, etc.
  - Precaution should not produce paralysis of science and technology
  - "Acceptable" risk is socialized (e.g. by means of insurances)

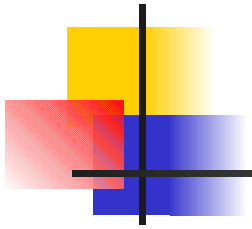
Deepening our understanding

Being cautious about precautionary principles



- Should one enforce a "human-in-the-control-loop" exceptionless requirement?

No! Machines *can* take decisions which humans should not override (e.g., to prevent accidents)



Ethically motivated research

- Improving machine learning standards
- Installing a “black box” into learning systems
- Providing machines with explanation & justification facilities
- Integrating learning and symbolic knowledge

Explanation and justification

- Accountability, autonomy, trust, social anxiety
- Machines should become increasingly capable to explain and justify their courses of action
  - Antecedents in knowledge-based decision support systems and expert systems
  - Future Developments: Machine introspective and reflective capacities