Moral Intelligence for Human and Artificial Agents

Laura Pană

University "Politehnica" of Bucharest
313, Splaiul Independentei,
Bucharest, Romania
lcpan20032000@yahoo.com

A specific form of human and artificial intelligence is needed for an efficient practice or a successful implementation of a moral code. The actual stage and rate of progress in AI studies are generated by their orientation to develop mainly an abstract intelligence form.

We propose the integrated analysis of the intelligence as of the morality, considered as systems with internal structure. Is prepared an artificial ethics, which is not an application of the human ethics, but an outcome of moral invention, and which can be equally practised by humans and machines.

Human and artificial beings are acting and interacting now in an artificial technical environment which is partially an intellectual environment. Machines cumulate multiple and meaningful functions related to man and society. Artificial agents are created not only for assist, but for replacing humans in processes as fabrication, business, services, communication, researches, education or entertainment. Their conduct receives thus a moral signification.

This more and more artificial world, generated by man-machine interaction, produces not just complication of the machine, but of the man himself and of its moral values. Human species evolves in all its dimensions: biotical, psychical, social and cultural: now it evolves towards artificiality.

Even as a biotical (natural) being, man evolves to artificiality: in our process of aging and decreasing, we receive more and more artificial components and we become more and more robots. Society evolves too, mainly under the influence of the information technology advances: we have now an information society or even a knowledge society.

The cultural dimension of humans is also transformed and a technical man is born. All types of values are renewed, because of the emergence of new human needs, often satisfied by virtual relations and virtual means, in a virtual environment. Moral values change their content, appear new values and the entire hierarchy of moral values is now modified. Starting from the fact that human and artificial agents are going now to explore and to populate a global virtual intellectual environment, we anticipate a movement from the ethics of the virtue to a virtual ethics.

Human morality tends to become more complex and hard practicable because of its diversity and relativity, being often reduced in practice to the professional deontology. Even past and present moral theories (ethical systems) present serious weaknesses, now analytically studied by computer ethics. Such conceptual problems, which can make difficult the work of moral norms as rules for machines, are not consequences of theoretical errors, but are generated even by the nature of human morality and its complexity.

Morality includes moral knowledge, conscience and action. Moral knowledge is an evaluative knowledge which evolves from philosophical to scientific and now, even towards technical knowledge. Moral conscience is highly structured but gradually formed and unequally developed, from habits, opinions, beliefs and convictions to moral reflections. Thus, morality is also a form of spiritual life: the understanding of specific values and the moral invention occur at this level.

All mentioned structural levels of human morality need for their manifestation not just intellectual skills, but constitution and use of some psychical aptitudes, personality traits and cultural orientations and attitudes.

Moral values are synthetic human values. They are present in all kinds of human action and any human action has a moral dimension. Dignity, which can be considered the central moral value, needs achievement of all other moral values. Each moral value needs, for its realization, one or more human aptitudes: intelligence is asked to exert the capacity to do what is right or good; imagination is used to avoid the possibility to proceed badly; consequence permits to obtain moral sincerity; attention, intuition and the unlimited aptitude to learn ensure moral responsibility etc.

Yet we must add that moral values and even norms are vague and changeable in content, often unclear in form and need concrete application in different domains and contexts of action. Interpretation of moral values for their transposition in moral norms and then in human conduct generates the moral conscience, as node of links but also as knot of conflicts between ideal values and real bearing.

Human morality is, also, a morality of preference: moral decision is based on individual or group interests, which

become criteria of decision; moral life is characterized by deep affective involvement and implies emotional intelligence; morality is credited as altruist conduct, but appears as mean calculus; this counting is a subjective calculus of probability; often theoretically wrong, human moral computing can be practically efficient, but is always declared immoral.

But such an imperfect morality needs perfect instruments for its implementation: applications of special logic fields as logic of preference or logic of belief and a precise and reliable technical logic. As spiritual life form, morality has a psychical infrastructure and can require studies in psychological techniques.

For the spiritual level of morality, efficient models can be structured and adaptive moral programs can be selected by computer simulation in a complex moral environment, inhabited by populations of moral agents, differentiated by their duties.

But because of their level of complexity and mainly by their necessary degree of freedom, artificial moral agents must be conceived and realized as individual entities, endowed, in addition, with other necessary qualities.

Artificial moral agents can/must be treated as 1 - individual entities (complex, specialized, autonomous or self-determined, even unpredictable ones), 2 - open and even free conduct performing systems (with specific, flexible and heuristic mechanisms and procedures of decision), 3 - cultural beings: the free conduct gives cultural value to the action of a human (natural) or artificial being, 4 - educable, not just teachable systems, 5 - entities with "lifegraphy", not just "stategraphy", 6 - endowed with diverse or even multiple intelligence forms, like moral intelligence, 7 - equipped not just with automatisms and intelligence, but with beliefs (cognitive, evaluative and affective complexes), 8 - capable even of reflection (moral life is a form of spiritual, not just of conscious activity), 9 - components/members of some real (corporal or virtual) community.

Implementation of such characteristics does not necessarily suppose efforts to design, construct and educate machine as quasi-human being. On the other hand, human conduct is perfectible just by the construction and application of a philosophically founded and scientific deduced ethics. This will be an artificial ethics, which may be applied by humans and machines, who can meet in the middle of the road between the natural and the artificial, in a common, better ethics.

Human intelligence is diversified as human activity. Our hypothesis regarding the moral intelligence existence and functionality is allowed by the application of systemic methodology in psychology and is sustained by an integrative philosophical vision about the forms of culture.

Researches concerning the structure of human intelligence evolved in two directions, but each of them left an incomplete map of its representation. By an inductive way, the mathematical, linguistic, descriptive, interpretative or theoretical forms of intelligence were studied, but not the scientific intelligence. The literary, musical or plastic intelligence were inventoried but not the artistic one. By the deductive way, a general intellectual functional availability (the so called general intelligence) was identified, and then two semi-general factors, responsible for an abstract intelligence form, respective for a more practical form. Later some special factors were studied, but these were considered common, in different proportions, for all forms of intelligence.

Technical intelligence was analyzed as a form of practical intelligence. Moral and political intelligence are also preponderant action oriented and strongly controlled by norms, but differentiated by the specific values pursued and by the kind of means utilized. Moral intelligence however cannot be integrated without difficulties in the group of practical intelligence forms, because, as we already found, achievement of moral values suppose moral knowledge, moral conscience and a complex spiritual life.

We have thus a general intelligence which ensures the specific, intellectual level of any human conduct, then particular forms of intelligence which include abstract and practical types of intelligence and finally, special forms of intelligence, generated by distinct domains of actions, oriented by specific values and developed by educational technologies and by experience in adequate environments.

Machine ethics can overtake some difficulties of human morality because it can be:

a - directly deduced from moral theory, *b* - assisted by intellectual techniques, *c* - based on objective evaluation of possibilities related to necessities, *d* - implemented by technical means which assure precision, transparence and efficiency, *e* - achieved by knowledge based technologies and cognitive robotics. But belief generating or conditioning and controlling problems will appear at the spiritual level of the artificial moral conscience.

We have analyzed three types of hypotheses for the effective implementation of a moral code, adequate for artificial moral agents: a structural, a functional and a behavioral hypothesis. These hypotheses are based on the postulate that complexity of such agents needs to allow them freedom, not just to expect responsibility from them. Developed in a recent paper, these hypotheses regard the implementation of a new, entirely invented moral code, not a tentative to transpose a human ethics in a machine ethics. Both the original and the model are artificial.

Artificial ethics, as a solution to avoid both conceptual and technical difficulties of a moral code implementation, can also be considered as a new step of a process of generalization of artificiality in the human culture, from the techno-science, information aesthetics, digital politics and computer ethics to the artificial philosophy.