

humaine

# Ethics and human-machine interaction - a scaffold for thinking about ethics and human-machine interaction



Shlomo Hareli  
University of Haifa

WP10 workshop – Vienna

## Road map of the talk

- Setting the stage – the main questions
- Possible answers
- Analyzing an example
- Conclusions and further questions

## What is that context ?

- Humaine is about human-machine interaction.
- Computer-based systems that require people to interact with them in a manner which simulates aspects of social interactions between human beings.



## What do we mean when we consider ethical questions in relation with HUMAINE?

- Ethics in this context is about the “right and wrong” in human-machine-interaction.
- Chiefly about what we are allowed to do as system developers, testers and users and what not. What, as a consequence, a system is “allowed” to do or not. All this taking into consideration that the effects of a given system on a user can be both direct and indirect and can occur during an interaction with the system but also prior to it or long after it ended.

For example, the system can cause damage and this damage can be either:

- Prior the interaction - one is computer illiterate and the mere knowledge that he would have to interact with a computer threatens him and makes him feel uncomfortable.
- During the interaction – an artificial agent is programmed to deceive the person who interacts with it in an attempt to persuade him to buy a new product.
- After the interaction – the system stores personal and confidential information about individuals interacting with it and that information is accessible to everyone.

And

- Potential subjects for the effects of the system are individuals who interact with a system or other people who may be somehow influenced by such an interaction. For example, a patient of a doctor who uses the system as part of the treatment that he offers to that patient.

## What is our ideal goal – the fantasy ?

- We may aspire to achieve a situation in which we are fully aware of the entire set of issues that impinge on the “right and wrong” in human-machine interaction.
- This is obviously unrealistic.

- There are endless types of applications that can be included under the heading of “human-machine interaction system”.
- There is an infinite number of issues associated with human-machine interaction systems and that give rise to ethical questions.
- Our ability to envision all the kinds of uses and abuses of a given system is limited. We fail to see what our creation can do apart from what it was planned to do. In this regard we are like God who regretted the creation of man as this creation went out of control. This is a great challenge to any complete ethical consideration.

So what can we do? The realistic outlook

- We can think of general principles, guiding rules that will guide our thinking about the ethical questions surrounding our systems.
- General markers that will serve as a scaffold for the building of specific ethical principles for particular types of systems.

### A preliminary scaffold

- The idea is to offer a scaffold incorporating some general guiding principles for suggesting ethical rules for automated systems interacting with human beings.
- At the heart of the scaffold is the idea that on a large scale, two major types of ethical issues will always be relevant in each case and for each system:

- (a) Ethical issues specific to the purpose of the system (e.g., an automated persuasion system or a system that performs automated medical intake interviews).
- (b) Ethical issues that are general and relevant to any automated system.

Ethical issues specific to a given type of system

- The “right and wrong” of the specific domain of the application.
- It is expected that the ethical rules existing in a specific domain will apply also for automated agents or other systems that perform tasks or serve goals similar to these performed in more conventional ways in the same field.
- For example, there is no reason to assume that an automated system that serves the goal of persuasion via advertisements will not have to comply with ethical rules accepted in that field.

- Of course, this will be the minimum ethical standards and they will not be sufficient.
- However, the system and its users must comply with such rules.
- The “good news” about that is that one doesn’t have to “invent the wheel”. There are “out there” ethical guidelines which are relevant and even mandatory for a given system. Probably, most applications mimic or resemble something for which someone (more than one) devised ethical rules.

### Example – an automated persuasion system used for advertisement

- Do general ethical rules devised for “conventional” advertisement apply for an automated system serving a similar goal?
  - YES
  - The American Advertising Federation (AAF).
  - Advertising Ethics and Principles
1. Truth  
Advertising shall tell the truth, and shall reveal significant facts, the omission of which would mislead the public .

## **2. Substantiation**

Advertising claims shall be substantiated by evidence in possession of the advertiser and advertising agency, prior to making such claims.

## **3. Comparisons**

Advertising shall refrain from making false, misleading, or unsubstantiated statements or claims about a competitor or his/her products or services.

## **3. Bait Advertising**

Advertising shall not offer products or services for sale unless such offer constitutes a bona fide effort to sell the advertising products or services and is not a device to switch consumers to other goods or services, usually higher priced .

## **4. Guarantees and Warranties**

Advertising of guarantees and warranties shall be explicit, with sufficient information to apprise consumers of their principal terms and limitations or, when space or time restrictions preclude such disclosures, the advertisement should clearly reveal where the full text of the guarantee or warranty can be examined before purchase.

## **6. Price Claims**

Advertising shall avoid price claims which are false or misleading, or saving claims which do not offer provable savings.

## **7. Testimonials**

Advertising containing testimonials shall be limited to those of competent witnesses who are reflecting a real and honest opinion or experience.

## **8. Taste And Decency**

Advertising shall be free of statements, illustrations or implications which are offensive to good taste or public decency.

\*Adopted by the American Advertising Federation Board of Directors, March 2, 1984, San Antonio, Texas.

All of these are significant also for automated persuasion system and must be adopted. Thus, a good source for these domain specific ethical rules are rules devised in the respective fields. At least, as a starting point.

Ethical issues that are general and relevant to any automated system.

- The “right and wrong” of the general domain of human-machine interaction systems.
- It is expected that some ethical rules that will be devised for systems that interact with human beings will share many features and hence one can think of a set of ethical issues that apply to all such systems.
- For example, there is no reason to assume that an automated system that serves the goal of persuasion via advertisements will not have to deal with questions concerning information security, such as protection of user’s privacy by assuring that any information provided by a user will not be public unless the user agrees to that.

- Other things to worry about are questions of equal accessibility to different populations, “damage free” interaction, protection of other rights of the user etc.

## Conclusion

- There are different sources to extract from the kind of issues that are relevant for a consideration of ethical rules for automated systems performing interactions with human beings.
- One has to map first the terrain of facets which are relevant to system in question.
- Two general facets will always be relevant; one general and the other specific.
- Other aspects can also emerge as important, e.g. the issue of testing a system. In this case too one can find relevant domains which already created and adopted relevant guidelines (e.g., experimental psychology).

## Remaining questions

- What will happen with systems for which a parallel non automated system does not exist?
- Can't there be an overlap or even contradiction between different ethical rules applying to a given system that come from considerations of different facets related to the system?
- If the above is possible, how we resolve such a problem?

THANKS FOR LISTENING