

# humaine

**D10c**  
**An Ethical Framework for**  
**Emotion-Related Computing**

*Workpackage 10 & 11 Deliverable*



**Date: 30<sup>th</sup> November 2006**



<b>IST project contract no.</b>	507422
<b>Project title</b>	<b>HUMAINE</b> <b>Human-Machine Interaction Network on Emotions</b>
<b>Contractual date of delivery</b>	<i>November 30, 2006</i>
<b>Actual date of delivery</b>	<i>December 6, 2006</i>
<b>Deliverable number</b>	D10c
<b>Deliverable title</b>	Ethics and Best Practice
<b>Type</b>	Report
<b>Number of pages</b>	20
<b>WP contributing to the deliverable</b>	WP10, WP11 and others groups
<b>Task leader</b>	University of Manchester
<b>Author(s)</b>	Sheelagh McGuinness, Peter Goldie, Ian Sneddon, Paolo Petta
<b>EC Project Officer</b>	Philippe Gelin

Address of lead author: Sheelagh McGuinness  
 School of Law  
 University of Manchester  
 Manchester  
 UK

---

## Table of Contents

<b>1</b>	<b>STATUS OF THIS REPORT .....</b>	<b>5</b>
<b>2</b>	<b>THEORETICAL ISSUES: PRINCIPLISM REVISITED.....</b>	<b>6</b>
2.1	A SHORT INTRODUCTION TO PRINCIPLISM.....	6
2.2	CRITICISMS OF PRINCIPLISM .....	7
2.3	PRINCIPLISM IN HUMAINE.....	8
<b>3</b>	<b>REPORT OF THE ETHICS COMMITTEE .....</b>	<b>10</b>
<b>4</b>	<b>APPLICATION OF ETHICS.....</b>	<b>12</b>
4.1	SOME THEORETICAL ISSUES.....	12
4.2	APPLICATION OF ETHICS TO CURRENT RESEARCH WITHIN HUMAINE .....	15
4.3	CONCLUSIONS .....	18
<b>5</b>	<b>OTHER RESEARCH PRESENTED AT THE WP10 WORKSHOP .....</b>	<b>19</b>
<b>6</b>	<b>CONCLUSION AND OUTLOOK.....</b>	<b>19</b>
	<b>BIBLIOGRAPHY .....</b>	<b>20</b>

# 1 Status of this Report

The aim of this report is to provide a summary of the in-depth discussions of ethical issues of the development of emotion-oriented systems that took place at the WP10 workshop in Vienna, November 6-8, 2006. Particular attention is paid to the role of principlism, as a primary method of ethical analysis. Furthermore, this deliverable also gives an account of best-practice procedures in the work on HUMAINE Exemplars. It closes with an assessment of the realities of integration of Ethics into the overall programme of activities of HUMAINE

## 2 Theoretical Issues: Principlism Revisited

### 2.1 A short introduction to Principlism

Principlism has been adopted as the ethical method by HUMAINE. (HUMAINE deliverable D10b) Some of the criticisms of the theory are worth consideration as it is important to be aware of the weaknesses and the strengths. The four principles approach was first championed by Beauchamp and Childress in the 1970s. (Beauchamp & Childress, 1979) Also referred to as ‘principlism’ or the ‘Georgetown Mantra’, after the institute to which the pair belonged at the time, it has become a standard approach to ethical decision-making among health-care professionals. Beauchamp and Childress see the four principles (=“4P”) of *beneficence*, *nonmaleficence*, *autonomy* and *justice* as norms in a universal, common morality. By this they mean that the principles are so fundamental that they will appeal to “all persons in all cultures who are serious about moral conduct”. Beneficence implies an obligation to do good for your patient and nonmaleficence a duty to do no harm. Autonomy implies a duty of non-interference, for example, respect for the decision-making capacity of an individual even if the consequences of these decisions are not in their best interests. To define justice is more problematic but at its most basic probably concerns access to health care and just distribution of healthcare resources.

They see the principles as giving rise to *prima facie* rather than *actual* obligations. (Beauchamp & Childress, 2001) *Prima facie* obligations can conflict with each other and a situation or decision can be both *prima facie* right and *prima facie* wrong at once. When this occurs these obligations must be further analyzed and weighted against each other to ascertain what is the ‘greatest balance’ between right and wrong. Throughout their account of principlism Beauchamp & Childress stress that the approach is not to be taken as a general moral theory. They recognize that alone the four principles are abstract and have insufficient content to deal with most problems in ethical reasoning. They see the principles as providing a framework and the processes of specifying and balancing build on this framework to move from principles to rules and ultimately decisions.

*Specification* plays a central role in principlism. While the principles provide a basis of values, specification removes a layer of abstraction and applies these values to specific issues that give rise to general rules. Specification provides “action-guiding content” to principlism. However it is not enough to create general rules; reasons and justification must also be provided in specification. This is a rigorous and dynamic process. However, the process of specification does not on its own solve conflicts that may arise between principles, as conflicts may also arise between specified rules within a single principle, eg: between competing demands of justice. Indeed the process of specification requires the individual to make different judgments about different principles without always giving guidance for these judgments. While specification deals with the ‘range and scope’ of each rule and principle, the process of *balancing* deals with the ‘weight and strength’ these rules should be given in ethical decision-making. Balancing is an extension of specification and is not always necessary as some specified rules are taken as absolute. It is, however, important when rules come into conflict with each other and no over-riding interest is easily identifiable. When balancing one rule against the other, reasons must be given and

these reasons cannot merely be subjective intuitions for believing one rule more important than the other. They must be morally justifiable reasons. Beauchamp & Childress lay down eight conditions which should be met in the process of balancing. They believe that these conditions reduce the intuitive aspects of the balancing process. Reasons given in reaching a conclusion to an ethical problem must be justifiable. They distinguish between *good reasons* and *morally justifiable reasons*; the latter are necessary to reach a morally justifiable decision.

The processes of specification and balancing will not always give one answer to an ethical problem, but Beauchamp & Childress do not see this as being a weakness in the four principles approach. They also acknowledge that *things other than principles* should be taken into account in decision-making, for example, ‘cultural expectations’ and ‘precedent’, all these factors combine to give weight to one particular outcome over another. In this final process of weighing one particular outcome against the other Beauchamp & Childress use a method similar to that of Rawls; ‘reflective equilibrium’. This process helps ensure that all the principles and rules used in reaching a decision are as coherent as possible. For Beauchamp & Childress ethical theories should be modified and revised in a manner similar to that of scientific hypotheses.

## 2.2 Criticisms of Principlism

Here are some of the criticisms which have been leveled against Principlism. Takala has questioned whether Beauchamp and Childress are justified in their claim of a common moral language. (Takala, 2001) She highlights that each principle will be subject to massively different interpretations depending on the context of what society they are being considered in. This would mean that although the *titles* of the four principles would remain universal, their *content* would be hugely varied. This highlights the fact that only the titles of the principles are universal and their interpretation across cultures and within societies can be diverse. Holm raises similar criticisms about the difficulties of trying to create a universal ethic. (Holm, 2001) This relativity seems to weaken Beauchamp and Childress’ claim of a common morality. Holm states that this could be overcome by letting the four principles point to important aspects of morality across cultures but leaving the exact content to be decided within each culture.

Clouser and Gert focus on this weakness in their criticisms of the four principles. (Clouser & Gert, 1990) They have leveled three main objections to ‘principlism’. The first of these is the criticism that the four principles are analogous to a checklist or a chapter heading but that they contain no moral substance. The four principles themselves seem to consist in nothing other than pointers, which may highlight considerations for ethical decision-making but provide little guidance in coming to a decision. Beauchamp & Childress openly concede this. They say that it was never their intention that the four principles alone would be sufficient to deal with ethical dilemmas or problems.

In light of the above criticisms it would seem that the principles are little more than the same chapter headings with assorted moral contents in different cultures and societies. Beauchamp & Childress are overly optimistic in their claim to a global ethic. Clouser and Gert’s second criticism is that the four principles provide no guidance to

an individual when confronted with a moral problem. This lack of guidance leaves the individual free to give a subjective account of what is to be considered. They can focus on whatever principle they believe to be most important and weight the corresponding rules accordingly. This leads to Clouser & Gert's third criticism, which is when two principles come into conflict, principlism is too indeterminate to give an account for ethical decision making. In reply to this objection Beauchamp & Childress point to the processes of specification and balancing as guiding processes. They believe that these processes give structure to the process of ethical reasoning.

Takala and Holm point to considerable problems with these processes. The obligations that the four principles give rise to, as I have mentioned earlier, are *prima facie* and this means that they can be discarded if there is good moral reason to do so. However as Takala points out, "even this prohibition against breaking a promise turns out to be empty as a universal principle". What will be considered a morally good reason will be subject to the same relativism in interpretation as the principles are themselves. Holm criticizes the lack of technique involved in the processes of balancing and specification themselves. He claims that Beauchamp and Childress give little criteria as to what should be considered morally relevant in our ethical deliberation.

### ***2.3 Principlism in HUMAINE***

On its own the approach appears to be of limited value. It is therefore not surprising that supporters stress the fact that it is complementary to and can be used alongside other general moral theories. On Beauchamp & Childress' account, casuistry and accounts of virtues seem to be the favored bedfellows for principlism. Considerable space is used to highlight the compatibility in *Principles of Biomedical Ethics*. Gillon even believes that principlism has the capacity to bridge the gap between utilitarianism and Kantianism: ethical theories that are generally thought to be in direct conflict with one another.

As a method it may still have much to offer. It highlights four important principles which are worth considering in ethical analysis, it has many benefits. Time and again it is stated by those involved in clinical ethics committees that the 4P provide a useful starting point. Many of the criticisms mentioned relate to the fact that medical treatment is often as cultural as it is scientific and principlism will take on different forms in different cultures—whether these criticisms will be true of the work being carried out by HUMAINE is not yet clear. But it is important when considering any methodological approach to be aware of the perceived downfalls as well as the merits of the approach.

When it appears that the Principles conflict with each other or are mutually incompatible it will be necessary to consider background conditions. Situations should always be considered in the appropriate context, which is not to say that we should give way to relativist accounts of 'anything goes'. These principles act merely as 'rules of thumb' and it may not always be necessary to consider all four in every situation. They do however provide a clear framework that can be used methodologically.

So why have we adopted principlism in HUMAINE? In spite of all the criticisms outlined above, principlism has many attractions, primarily the fact that as a method it is just intended to encapsulate our common sense morality. Common sense morality is not aiming to deliver up answers algorithmically and so in that sense the criticism of principlism is misplaced. The theoretical approach of HUMAINE is that of common sense morality as embodied in principlism. Within the principles we find simply an account of every day common sense that can be used to achieve the best outcome to any problems that may arise. They do not consist of abstract and vague ideas but rather a common sense morality that is properly informed by scientific and philosophical knowledge. This will help the researchers involved in this project to think sensibly in terms of ethics—something that it can be seen has been achieved so far. (cf. HUMAINE deliverable D10b)

### 3 Report of the Ethics Committee

There are many issues worth considering when analysing the role of ethics committees and ethical guidelines. Most, if not all, funding bodies and research institutes require that research must pass certain processes in order to gain ‘ethics approval’. In creating ethical guidelines substantive ethical consideration will be involved and then hopefully these guidelines will be useful to those in that field. However it seems less clear that when ethical consideration gives rise to a sort of ‘test’ where one passes or fails that this substantive consideration is still being carried out. Yes it may be true that the processes and consideration given to these ‘application forms’ are truly ethical, but it seems less clear whether the outcomes of their application can be said to be ethical in the same sense. Also it is important that ethical consideration not go down the route of political correctness and its use be over stretched. We should not concern ourselves with being so ethical as to lose sight of the original aims of ethics and we should avoid it being used in a tick box sense merely to provide ‘window dressing’ in order to get funding. It would seem that HUMAINE is in a good position to expand on this apparently arbitrary approach to ethical committees.

The HUMAINE Ethics Committee has as its aim to function in a way that can meet the needs of the HUMAINE research community. (Sneddon, 2006) This means that it should be a resource which members of that community feel they can access when they have ethical concerns about their research. It is important that researchers should be able to do this without fear that sanctions or prohibitions will be placed on their work. In order to fulfil this role the needs of the community must be identified, this presents an interesting challenge given the wide range of disciplines that work across the network.

The Ethical Audit carried out by Goldie & Döring in 2004 showed that many of the researchers within the community did not have access to an ethics committee. They also found that many researchers also were carrying out research that related to human subjects. The committee was set up to deal with these issues. Its purpose is to answer questions on research ethics and offer advice on procedures and best practice. It also has as an aim to review any HUMAINE research involving human participants not already scrutinised by an ethics committee. In this audit 27 partners were contacted of which 24 replied and details of the report can be found in HUMAINE deliverable D0f.

In 2006 the Ethics Committee circulated a very short questionnaire (four questions) asking researchers about present and future research projects and their access to an ethics committee. They received 29 replies from 23 partners. These were the findings:

- 13 (45%) were conducting research involving humans;
- Of these 13, 7 do **not** have access to an ethics committee;
- Of these 7, 5 intend to submit or have already submitted applications to us.

The Ethics Committee has received two formal applications which were decided not to raise any ethical issues. A third has been submitted for discussion on an informal basis. The remaining two cases are still pending and the committee awaits the applications, but informal discussions have taken place. The two applications where at present there is no declared intention to submit are being followed up also.

In the 2004 Audit it was shown that although 88% of respondents were conducting research involving human participants, over 80% of these researchers thought their research posed no risk to participants. While these researchers may be correct in their assessment this presumption should not be the end of the story. The research committee has as its aim the promotion of an increasing acceptance of the principle that the researcher should not be the sole judge of whether their research poses ethical problems.

There are many methods of promoting this. Firstly it must be constantly stressed that the committee is a resource for all researchers. The documentation required from ethics committees can often appear fussy and bureaucratic—HUMAINE must take steps to avoid this. It is also important that anyone assessing research gets consistent and appropriate information. This means that a clear set of ethical guidelines should be developed for the community. The chosen method of principlism can provide the structure for these guidelines. The content of the principles should be explained, using principlism, in the guidelines along with concrete examples of how they apply to the work of HUMAINE. This is not to say that ethics guidelines can be applied in an algorithmic way. When the context of different situations is analysed it will not always be appropriate to apply the principles arbitrarily. This is not to say that every situation that requires unique considerations will be creating ‘novel’ ethical concerns. Even if the science is new the ethical issues will probably not be.

The ethical guidelines can be a tool outlasting the HUMAINE project. It has also been suggested that it might be very useful to have a HUMAINE Committee which carries on after the project to deal with the administration and ongoing promotion of this and other products of the project. Indeed all the hard work can be seen to have paid off in the manner in which the ethics committee interacts with researchers in the project. This is something which is presently under consideration.

## 4 Application of Ethics

### 4.1 *Some Theoretical Issues*

When considering how best to apply ethics to the science and technology which HUMAINE is developing it is first necessary to look at some of the theoretical issues raised. The first of these is the question of who or what are we considering ethics for? Also there are more general obligations that we should be aware of, eg: issues of social responsibility. (Tzafrir, 2006) It is important to remember the principles of beneficence and nonmaleficence—we must be careful to be aware of the potential consequences of our research and the uses it may be put to.

When we consider ethics in HUMAINE who are we referring to? This is a many stranded question. On the one hand there are the ethical concerns which engineers and scientists should be aware of as they carry out their research. Then there are also the implications of this research. Is the aim to make autonomous agents? What standards of autonomy would have to be reached in order to achieve this? If it is not possible to make actually autonomous agents but rather agents that appear to the user to be autonomous then different concerns will be at issue. These concerns will be heightened when these apparently autonomous agents are ‘trying’ to persuade the agent to act in a certain fashion. (Guerini & Stock, 2006)

If it is the aim of HUMAINE research to create an autonomous agent it will be necessary to have a working standard of autonomy. Baumann & Döring suggest that the following procedural conditions of autonomy would have to be reached: self-reflection, rationality, procedural independence, self-trust/ -esteem/ -worth. (Baumann & Döring, 2006) All these conditions they suggest must be fulfilled for an agent to be considered autonomous. Self-reflection will involve conditions of self-awareness, self-understanding and self-evaluation. Rationality will involve means-ends-reasoning, rational coherence and critical reasoning. The final considerations of self-trust, self-esteem and self-worth are essential in order to exercise its capabilities for self-reflection and rationality. These conditions are socially constituted feelings, which means that they are dependant upon the recognition of others. Not only must an agent fulfil these procedural requirements but also it must reach them in a procedurally independent way. It is unlikely that a machine would ever be able to reach these standards. We are not dealing with science fiction and it is important that this point be remembered when we deal with hypothetical ethical issues.

When creating interactive machines, there must be an awareness of the ethical issues that will arise between user and machine. When machines are not themselves capable of ethical reasoning, many of the responsibilities will lie with the programmer. They must consider what it is they wish their machine to do and what possible effects this may have on users. It is also important to be aware of how the machines could be misused by others in the future, for example machines with the capacity to deceive the user may be able to cause the user much harm. If it is a persuasive agent, the machine may in a sense challenge the user’s autonomy. However while it is important to be aware of these things one must not be overly aware of them. Whether human machine interaction can create genuinely novel situations is unlikely and the principles of beneficence, nonmalficence, autonomy and justice will provide a solid

framework within which different possible outcomes can be examined. It may also be useful to divide ethical considerations in two groups—those that relate specifically to the purpose of the machine and then more general issues with automated systems. When considering the uses of automated machines it is important to remember the general guidelines that apply in that field, for example when creating persuasive machines for use in the advertisement industry there may be advertising standards or regulation of advertising methods which must be abided by. (Hareli, 2006) This again goes to the heart of whether the issues related by human machine interaction can be described as novel.

Issues of user autonomy may provide different kinds of concerns. If we are creating machines which appear to be autonomous and with which the user is interacting, it will be necessary to consider how best to protect the autonomy of the user. When we design ECAs or Avatars that specifically influence or effect user autonomy then there must be clear guidance as to who is responsible for this impact on autonomy. If a machine is created in order to persuade the user to act against their own interests, is this different from another human being doing the same thing? It must also be remembered that so long as the user maintains the capacity to reason and make choices for themselves they will be responsible for the choices that they make. Here are some practical examples of this (Baumann & Döring, 2006):

<p><b>Hiding true intentions</b> An emotion oriented system does not inform the user that it will try to persuade her/him.</p> <p><i>Possible Threats</i> The person cannot become aware of his/her social environment and therefore lacks self-awareness. As a consequence, s/he is not able to adequately deal with the information provided to him/her and cannot engage in critical reasoning.</p> <p><i>Ethically permissible?</i> (from the perspective of autonomy) “It depends”: Should we rationally expect that the user knows or comes to know that s/he is being persuaded?</p>	<p><b>Hiding important information</b> An emotion oriented system hides important information</p> <p><i>Possible Threats</i> A person cannot engage in effective <i>means-end reasoning</i> because it is based on incomplete information. Dependent upon the kind of information: Detrimental to self-understanding or to self-evaluation.</p> <p><i>Ethically permissible?</i> “No” (but: depends upon the kind of information): it is unlikely that it can be expected from the user that s/he has access to these information.</p>
<p><b>Lying</b> An emotion oriented system lies to a user.</p> <p><i>Possible Threats</i> A person cannot engage in effective <i>means-end reasoning</i>, because his/her reasoning is based on false premises; the same holds for self-evaluation. A person's self-respect might be diminished – s/he might be confronted with information which do not cohere with his/her other beliefs and, in consequence, the person might doubt that s/he is a competent deliberator; especially when the false information is presented in certain authoritative ways.</p> <p><i>Ethically permissible?</i> “No”</p>	<p><b>Inducing Extreme Emotions</b> An emotion-oriented system induces extreme positive/ negative emotions.</p> <p><i>Possible Threats</i> A person's capabilities for <i>self-reflection</i> are <i>incapacitated</i>. A person is unable to exhibit rational control.</p> <p><i>Ethically Permissible</i> “No”: It does not seem reasonable to expect from the user that s/he be able to resist the influence of extreme emotions on these capabilities</p>

<p><b>Altering users' emotional state I</b> An emotion-oriented systems alters a person's emotional state subliminally</p> <p><i>Possible Threats</i> The person might not be aware of the influence the emotion has on his/her actions or decisions (lack of self-awareness) The person might lack self-understanding, since s/he cannot make sense of his/her emotional state - how s/he came to have it, why s/he has it</p> <p><i>Ethically permissible?</i> "It depends": Is it possible for the user to become aware of his/her emotion and of the processes that gave rise to it (making his/her emotional state intelligible to him-/herself)?</p>	<p><b>Altering users' emotional state II</b> An emotion-oriented systems alters a person's emotional state against his/her own free choice</p> <p><i>Possible Threats</i> This is not a case of persuasion. It is plain coercion. A person is no longer able to exercise control over his/her life.</p> <p><i>Ethically Permissible</i> "No"</p>
--	--

What of the machines—what responsibility can a machine have? And is it even possible to create machines with the capacity to reason ethically? For instance would it be possible to have a machine capable of solving ethical dilemmas? In order to achieve this goal it would be necessary to be able to create an algorithmic programme for ethical evaluation—whether this is possible seems at present unlikely. What may be possible is to program the computer in a way that makes it act according to certain rules and give it the further guidance on what to do in situations where these rules clash. This however does not seem to be ethical deliberation in any real sense since the computer lacks the ability to understand the choices that it makes in these situations. Another point about programming computers to act in an ethical manner is that this will involve a consideration of the best possible outcomes; this totally disregards the role of intention in ethical deliberation. Therefore it seems unlikely that it would be possible to create a machine that can act as a moral agent; ultimately the machine would need to be controlled by a human being. However what is possible is the creation of a machine with the attributes of being an ethical agent and this will need to be considered using the guidance set out above.

Finally it is worth considering again some of the responsibilities that the designer or programmer may have when creating machines that either have the capacity to act autonomously or have the attributes of acting in this manner. The following checklist of questions should be kept in mind:

- Will this application be capable of lying? If so what is the purpose?
- Will this application be collecting information about the user?
- Will this application be capable of deceiving the user?
- Will this application be capable of inducing emotion in the user?
- Will this application be capable of persuading the user? If so using what methods?
- Will this application be acting in the users interests?
- Will this application be capable to responding to emotion in the user? If so how and why?
- Will this application be capable of recognising emotional/ mental states of the user? If so how and why?

(Guerini & Stock, 2006)

With all of these questions guidance may often be found in pre-existing guidelines that exist with regard to the purposes for which the application is created, e.g.,

advertising standards. When programming a machine with certain ethical characteristics it is worth considering where the definition of that emotion has come from, what emotions you are choosing to use and those that are being left out. How will all these choices effect the continuing view of certain emotions and the marginalisation of those that are left out–will it be possible to create new emotions and for others to cease to exist? How emotions are defined will rely on a clear theory of emotion to provide a framework for programming choices. (Sussmann, 2006)

## 4.2 *Application of Ethics to current Research within HUMAINE*

When individuals interact with machines certain things are evident. If the machine appears socially intelligent then the user tends to treat it as human, e.g.: trust it/ believe it; however users also carry on treating them as computers, e.g.: turning them off. (Pelachaud, 2006) This shows that our perception of machines is often unclear. At once they are anthropomorphised and at the same time they are considered as machines. What is it we think the machine is? Should it only ever be a tool for the human to use? The following examples show how ethical considerations can be applied to the applications that are currently being developed in HUMAINE. This research was presented at the WP10 workshop in Vienna; presentation slides and background material are available on the HUMAINE portal at <http://emotion-research.net/ws/wp10/>.

### **\*Think Aloud\***

This is a study being carried out by Mina Vasalou, Intelligent Systems and Networks Group, Electrical and Electronic Engineering, Imperial College London, UK, and Tanja Bänziger, Geneva Emotion Research Group, University of Geneva.

- What is think aloud? A verbalisation method of ‘saying what you are doing’.
- What was the purpose of this study? Using the think aloud method as a concurrent method for evaluation in the lab.
  - Does the think aloud method allow users in the lab to be expressive?
  - Can a usability experts draw information about the user’s experience (along with the performance metrics often collected)?

This study involved 20 users to take part in the following scenarios:

- Rehearse a first time date while using Yahoo! Avatars
- Surprise your mom with a Yahoo! Card announcing your joint holiday in a Greek island

The users described their actions as they carried out these activities and described any user problems they encountered.

<b>Principle</b>	<b>Justice</b>	<b>Beneficence</b>	<b>Nonmalificence</b>	<b>Autonomy</b>
<b>Issues to consider</b>	There do not appear to be any obvious issues of justice arising from the present study.	For what reasons is information about the user being gathered?	Does this increase participants’ self-awareness to their detriment?	By gathering this type of information, may future user autonomy be infringed on?

## The W3C Incubator Group

This is an activity being coordinated by Marc Schröder, DFKI, Germany.

- What is the W3C incubator group? “The W3C Incubator Activity fosters rapid development, on a time scale of **a year or less**, of **new** Web-related concepts. Target concepts include **innovative ideas** for specifications, guidelines, and applications that are **not (or not yet) clear candidates as Web standards** developed through the more thorough process afforded by the W3C Recommendation Track.”
- What is the purpose of the study? They study questions of whether Emotion Annotation and Representation Language (EARL) and others fields are mature enough to be used as a web standard.

The study aims to investigate the prospectus of defining a general-purpose Emotion annotation and representation language which should be usable in a large variety of technological contexts where emotions need to be represented.

Principle	Justice	Beneficence	Nonmalificence	Autonomy
<b>Issues to consider</b>	The World Wide Web and its uses raises many questions of justice which may be of interest to this study group.	This study does not seem to raise any apparent questions about beneficence.	This study does not seem to raise any apparent questions about nonmalificence.	This study does not seem to raise any apparent questions about autonomy.

## Empowerment

This is a study being carried out by Kristina Höök and Jarmo Laaksolahti, Department of Computer and Systems Sciences, Stockholm University/Royal Institute of Technology.

- What is empowerment? This is a strategy which aims at dealing with human values in affective interactive systems.
- What is the aim of the strategy? To put users in charge of sense-making and definition of emotion.

This study hopes to protect user’s privacy and autonomy by making them privileged interpreters and constructors of content of an affective diary. This diary was created using a biosensor armband and a mobile phone. The diary had an open ‘surface’ where users could inscribe their own interpretation of what the phone and sensor showed. They could also fill the ‘surface’ with their own content.

Principle	Justice	Beneficence	Nonmalificence	Autonomy
<b>Issues to consider</b>	There do not appear to be any obvious issues of justice arising from the present study.	The users' privacy is protected by handing over the power tools to the end-users to negotiate what they want to share with one- another.	The users could tinker with and alter the emotional representations over time, this means that they would be able to protect their own interests over time.	Because the user is the privileged interpreter and constructor of the information in the diary this to a certain extent protects their autonomy.

### Validation of an affective model: A possible world model for emotional conflict

This is a joint study being carried out by Fiorella de Rosis and Peter Goldie.

- What is the purpose of this study? To create a model of emotional conflict.
- Three main questions were posed:
  - Why does it seem that emotion conflict is more common where the emotion is negative?
  - How can conflicts arise even when we are highly confident or even know that we are 'safe'?
  - How do we operationalise emotion- belief conflict?

Principle	Justice	Beneficence	Nonmalificence	Autonomy
<b>Issues to consider</b>	Who will be using the models of emotional conflict and for what purposes will the models be used?	What harms may causing a subject to remember negative emotions have?	Would it be possible to have a model which could 'cause' users to feel unsafe using this data?	How may user autonomy be infringed on by asking the user to recall particular emotions and then creating a model of these emotions?

### Abuses in Building and Using ECA Systems

This is a study being carried out by Catherine Pelachaud, LINC - University of Paris 8

- What was the purpose of this study? To examine the interaction between users and machines that can be described as socially intelligent.  
This study used 'believable agents' in order to study how users interact with socially intelligent machines. 'Believable agents' can be defined as 'a character is considered to be believable if it allows the audience to suspend their disbelief'.
- What were the aims of the study? To examine the various ways in which users and machines interact and also to examine the possibilities of abuses on user by the machine and on the machine by the user. It highlighted the fact that individuals have less concerns about abuse to machines than other humans.

The study concluded that one of the most important findings was the necessity to ensure that when the interaction between user and machine finished that the user be made aware that they were communicating with a virtual entity.

<b>Principle</b>	<b>Justice</b>	<b>Beneficence</b>	<b>Nonmalificence</b>	<b>Autonomy</b>
<b>Issues to consider</b>	There doesn't appear to be any obvious issues of justice arising from the present study.	In what way can socially intelligent computers manipulate an individual?	Can the user be damaged by the knowledge that they have abused the socially intelligent computer?	In what way is user autonomy affected by a computer with the ability to manipulate it/control the reactions of the user?

### **4.3 Conclusions**

The above is purely an account of the ethical issues that may be raised by human-machine interaction. It is in these situations primarily that new issues will arise. Although the situations may be new, it is not clear that this will give rise to novel ethical issues. As can be seen, a consideration of the general ethical principles will usually provide a framework in which to ensure that best ethical standards are met. There are other types of situations which may arise through human-machine interaction. These occur when the machine facilitates human-human interaction in some way, e.g., Internet relationships. (Ben- Ze'ev, 2006) However it is not clear that these issues will be much different to other types of human-human interaction, such as letter writing. Even so, in the realms of human-human interaction facilitated by machines the same principles of ethics should be considered.

These presentations in Vienna are clear evidence of the good work that is being done within the HUMAINE project to create an environment in which ethics and best scientific practice work together. The harmony between the research and the ethics is, we believe, a goal that all projects should strive for.

## 5 Other Research presented at the WP10 Workshop

There were other reports made at the WP10 workshop which do not fit under the previous heading of ethics and research.

The WP5 team reported back on the work that is being carried out on a pilot database exemplar. (Douglas-Cowie et al, 2006) At the minute, it is aiming towards a standard labelling scheme. The scheme should have a system that is relevant and contains a range of data, from induced to naturalistic. It will also contain samples from emotion in action and interaction. The scheme will describe the emotional content and the signs that convey it, and will also describe relevant contextual factors to interpret the signs. Also important will be the specification of the applications of the labelling scheme. The following modules have been proposed as part of the scheme: Emotion global module (static); emotion over time module; Speech and language module; Gesture/ face module; Physiological module. Future aims of the research will be to collaborate with EARL and also to address further some of the technical issues that have been raised.

## 6 Conclusion and Outlook

In HUMAINE, we are not dealing with science fiction, and this is one of the most important things to remember. When considering hypothetical situations we must distinguish between what is a real concern and that which will never be possible. Using the common sense morality of principlism we can deal with situations that may be new in a scientific sense but will not always be new in an ethical sense. Good work has been done to create an Ethics Committee that is taken seriously and fulfilling its role as an invaluable resource for researchers. The level of discussion in Vienna, as at other meetings, has been excellent and shows clear evidence of an awareness of ethical issues that extend beyond those considered by the Ethics Committee, e.g.: issues of the autonomy of Avatars and issues of social responsibility. The practices being put forward are not merely fulfilling the standards of what is required but aiming at higher standards of best practice. This type of interaction will much benefit the research that is being carried out by HUMAINE.

## Bibliography

Material related to the WP10 workshop is available on the HUMAINE portal at <http://emotion-research.net/ws/wp10>

Baumann, H. & Döring, S. 'Emotion Orientated Systems- Threats to User Autonomy' WP10 Workshop, Vienna, November 6-8, 2006

Beauchamp, Tom L. & Childress, James F. *Principles of Biomedical Ethics* [1<sup>st</sup> Edition] (New York: Oxford University Press, 1979)

Beauchamp, Tom L. & Childress, James F. *Principles of Biomedical Ethics* [5<sup>th</sup> Edition] (New York: Oxford University Press, 2001)

Ben-Ze'ev, A. 'Emotions and morality in computer-mediated and romantic relationships' WP10 Workshop, Vienna, November 6-8, 2006

Clouser, K. Danner & Gert, Bernard 'A critique of Principlism', *The Journal of Medicine and Philosophy* **15** (April, 1990) 219–236

Douglas-Cowie, E. 'The pilot database exemplar: Towards a standard labelling scheme' WP10 Workshop, Vienna, November 6-8, 2006

Guerini, M & Stock, O. 'Ethics Guidelines for Persuasive Systems' WP10 Workshop, Vienna, November 6-8, 2006

Hareli, S. 'Ethics and Human-machine interaction- a scaffold for thinking about ethics and human-machine interaction' WP10 Workshop, Vienna, November 6-8, 2006

Holm, S. 'Not Just Autonomy: The Principles of American Biomedical Ethics', in John Harris (ed.), *Bioethics* (Oxford: Oxford University Press, 2001)

HUMAINE deliverable D0f: Ethical Audit 2004

HUMAINE deliverable D10b Interim report to plenary meeting on ethical frameworks for emotion-orientated systems, 2005

Pelachaud, C. 'Abuses in Building and Using ECA Systems' WP10 Workshop, Vienna, November 6-8, 2006

Sneddon, I. 'The Humaine Ethics Committee' WP10 Workshop, Vienna, November 6-8, 2006

Sussmann, N. 'From (Artificial) Intelligence to (Artificial) Emotions', WP10 Workshop, Vienna, November 6-8, 2006

Takala, T. 'What Is Wrong with Global Bioethics? On the Limitations of the Four Principles Approach' *Cambridge Quarterly of Healthcare Ethics* **10** (2001) 72-77

Tzafir, S. 'Humaine- Research Network Social Responsibility' WP10 Workshop, Vienna, November 6-8, 2006