
Risks and Potentials of Affective Computing. Why the ACM Code of Ethics Requires a Substantial Revision

Thomas Grote

University of Tübingen

thomas.grote@uni-tuebingen.de

Oliver Korn

Offenburg University

oliver.korn@acm.org

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Verdana 7 point font. Please do not change the size of this text box.

Each submission will be assigned a unique DOI string to be included here.

Abstract

In this work, we discuss the ethical challenges arising from the transition of affective computing from the lab to the real world. Emotion recognition, especially from a distance and without body sensors, endangers central concepts of human autonomy and privacy. From both an ethical and an HCI perspective, we discuss how the ACM code of ethics (CODE) could be revised to better address such challenges. We suggest to employ positive law and to use “middle-range-principles” to build a substructure of the CODE that researchers in HCI can apply more easily. However, as positive law is applied by studying individual cases, a public discussion about ethical issues in HCI is required.

Author Keywords

Ethics of HCI; affective computing; moral principles; casuistry; usability; middle-range principles.

ACM Classification Keywords

K.4.1 Public Policy Issues: Abuse and crime involving computers, Ethics, Use/abuse of power; H.5.1 Multi-media Information Systems: Artificial, augmented, and virtual realities; H.5.2 User Interfaces: Interaction styles, User-centered design; J.4 Psychology, Sociology



Figure 1: A student wearing a tracking device for brainwaves.



Figure 2: A student's hand with a tracking device for skin conductance (GSR, galvanic skin response).

Introduction

Affective computing is an area in HCI developing sensor based computational systems which are able to interpret and simulate emotions. While it was established over 20 years ago by Picard [1], the current breakthroughs in system integration, machine learning and big data analysis allow powerful applications full of potentials – and threats. Affective computing has the potential to transform multiple areas in society, such as education, care and even the modern working environment – for better or for worse.

To illustrate this potential, we will give an example from the field of education: emotion recognition enables an interactive system to track a student's emotional state (Figures 1, 2) and in consequence, it might adapt in accordance to his or her (in future: female form) needs. On the other hand, this emotion recognition can infringe the student's privacy in situations where she is especially vulnerable. The system might use the emotional data to manipulate her; superiors could access delicate information on emotional stability and stress tolerance.

An even more problematic example is the detection of emotions from a distance, e.g. by facial expression analysis. With a good camera this can already be realized today, e.g. in public places like train stations.

Thus, it is evident that the technological progress in affective computing needs to be accompanied by careful ethical reflection about the values that the relevant technologies promote. Assuming this background, it also becomes evident that an ethical code of conduct guiding future research in affective computing is of critical importance.

However, in several cases the current Code of Ethics and Professional Conduct (CODE) of the Association of Computing Machinery (ACM) is unable to provide a satisfying normative framework for ethical decision making in the context of HCI research. One area where this becomes apparent is the implementation of a new technology into the social domain. Before we release a new technology "into the wild" the positive and negative effects need to be evaluated. Furthermore, we need to weigh these effects in a way that allows us to determine: is the direction this new technology is heading towards, worth pursuing? Is it morally right to step forward? More examples will be provided in the Related Work section.

Unfortunately, the current CODE does not provide HCI researchers with the tools to resolve those kind of ethical conflicts. There has been an ongoing debate whether and how the CODE needs to be revised [2, 3]. After nearly two decades, in which information technology has exponentially progressed and society has changed profoundly, we think it is about time.

We want to contribute to this debate by (1) identifying some of the CODE's structural weaknesses, and (2) suggesting how to develop a code of ethical and professional conduct which fits to the current needs of researchers in HCI. These suggestions are based on the idea that the ethical discourse in HCI research should emulate some basic principles of positive law.

In this work, "positive law" (to be distinguished from "natural law") stands for the idea, that legal rights are not inherent properties of certain groups, but human made and thus defeasible. As natural law theorists tend to spell out legal rules in a highly abstract manner,

positive law is typically applied by studying individual cases, which in turn become legally binding. Hence, we will try to sketch some ideas, about how such a case study-based approach can be applied in the ethical discourse in HCI.

Related Work

For HCI researchers ethical conflicts are rather common: often a technology that brings a significant positive effect to society also allows abuse, which could cause a proportionate amount of harm. Cyber warfare or even just cyber bullying [4] show that bad human habits transgress as neatly into the digital world as good ones.

Another example of such abuse: persuasive technology meant to increase their users' wellbeing has led to the emergence of dark patterns in web design [5]. Another challenge HCI-researchers have to deal with on a frequent basis are "value conflicts". Abstractly speaking, different values do not always fit neatly into a single normative framework. To give an example inspired by the contemporary societal climate: it is widely accepted, that personal freedom is a high good. The same goes for security. While these values are certainly not incompatible, they can under certain circumstances collide with each other. A government that virtually knows everything about its population could be in an ideal situation in providing security for them. On the other hand, this would certainly interfere with the citizens' personal freedom, as the government would be in possession of all their data.

In HCI, such value conflicts are an all too common phenomena. New technologies are often walking on a thin line between enhancing the abilities of the users

and undermining their autonomy by manipulating their behavior and decision-making. As a result, HCI researchers need to assess carefully the benefits and risks of the relevant technology. Furthermore, the researchers will sometimes find themselves in a position where they are forced to make a compromise between the technology's effectiveness and concerns for the wellbeing of the users by designing the technology in a way that is sensitive to moral values.

We illustrate this with another example from affective computing: a system gathering data on the emotional behavior over a longer period, could adapt extremely well to its users' needs. However, that system would also gather data creating a psychological profile of the users – based on multiple personal elements like working habits, stress levels, heartrate or attention span. If a superior or a health insurance company had access to this data, this would represent a major breach of privacy. Thus, in order to protect users from such a data abuse, a researcher needs to find ways to prevent the relevant data from being accessible to others, e.g. through black boxing the emotion recognition system as proposed by Korn [6].

Such a value-sensitive design is a demanding task, as it requires both technological competence and awareness of the relevant moral and legal issues and concepts. While we have good reasons to expect the former from a computer scientist, the latter is a more complicated issue: computer scientists usually receive little training in ethics or in legal matters. Thus, there is a need for methods and tools facilitating the ethical reflection on technological design. This line of reasoning should make clear, why ethical guidelines are of crucial importance for HCI-researchers.

Requirements

Before discussing some of the current CODE's weaknesses, we look at its general structure. It consists of 24 moral imperatives formulated as statements of professional responsibility. Moreover, the imperatives are divided into four sections. The first section entails the general moral imperatives, whereas the other sections refer to specific areas of professional conduct.

Through the generality of the first section's imperatives, each of them is formulated in a rather abstract way. For example, the first imperative emphasizes the responsibility of ACM members to contribute to society's wellbeing, while the second imperative emphasizes the duty not to harm others. The other imperatives in the first section refer to what one might call professional virtues, such as honesty, trustworthiness, fairness and the respect for the privacy of others.

As this list indicates, the CODE is based on a top-down approach in ethics, as a small set of general rules apply to a large range of potential ethical conflicts. Such an approach is to be distinguished from a bottom-up-model, such as casuistry. The latter is an inductive approach: its prime focus is studying paradigmatic cases of morally relevant issues; based on them, moral principles are construed [7, 8].

In principle, none of the CODE's general moral imperatives should spark much controversy: it is easy to see why promoting to society's wellbeing and preventing harm are desirable. In addition, honesty, trustworthiness and fairness are all laudable character traits. However, this already shows a major problem with the current version of the CODE: none of the key-

concepts of the relevant moral imperatives are well-defined. From the point of view of an ethicist, the meaning of concepts such as "harm", "fairness" or "privacy" is controversial. In fact, one might say that a huge debate revolves around those concepts in ethics. It is unlikely that HCI researchers have a deeper implicit understanding of these controversial concepts. This lack of semantic clarity is related to a second, more crucial issue with the current CODE: while all moral imperatives are abstract, they are not embedded into an overarching ethical theory. While some of the individual imperatives share some commonalities with utilitarianism or virtue ethics, the CODE is not committed to an ethical theory at large.

From an ethicist's viewpoint, this is problematic. Broadly speaking, we should expect at least two things from a code of ethical conduct: (1) it should pose a guideline for research that is in accordance with norms of morality and with legal norms and (2) in cases of moral conflicts, the code needs to be applicable in order to help resolving those conflicts.

This kind of "usability" is the key element, the current version of the CODE is lacking. When reflecting about the potential dangers and benefits of a new technology, we typically are confronted with different kinds of normative claims. Sometimes, those claims collide with each other. Emotion recognition in adaptive cognitive assistive systems for example might contribute to its user's wellbeing by reducing her stress-levels.

However, gathering intimate personal data like emotional stability is a threat to privacy. What should we weigh higher: a student's wellbeing or her privacy? While one should not confuse a code of ethical conduct

with an ethical theory, the CODE is very quiet on how to cope with these practical problems. However, the example of a moral conflict we described is no hypothetical scenario, but a current issue, when integrating emotion recognition in education [9]. In consequence, we advise a substantial revision of the CODE.

Concept for a Revised CODE

How would a suitable alternative for the CODE look like? We think that at least two structural changes are required. Both changes do not affect the moral imperatives per se, as these strike us as adequate. Instead, we aim for methods and tools that work as a substructure of the CODE, focusing on its usability or respectively, the operationalization its moral imperatives.

These can then serve as a basis for individual cases, which can become legally binding. One way to achieve this is implementing so-called "middle-range-principles" [10] into the CODE.

The concept of middle-range originates from the work of Beauchamp and Childress in bioethics [11]. The underlying idea is that there is a substantial distinction between normative ethics on the one hand and applied ethics on the other. The distinction is motivated by reasons that have much in common with our line of thought. Broadly put, the prospect of normative ethics is to develop a theory of morality for an ideal world. As the real world is quite "messy", if we want to determine what to do in a practical situation, we need "principles that are practically useful for analyzing the real world" [10]. This is where middle-range principles come into play.

Concrete examples are the "cost-benefit-principle", the "precautionary-principle", "sustainability", "autonomy" and "fairness". All these principles are arguably still rather abstract and there is an ongoing philosophical debate about their technicalities. However, they also have something in common: each principle has a rich history of application in different fields of society and, more importantly, in engineering ethics or technological assessment. This is especially true for the first three principles. In consequence, researchers in HCI should find it much easier to apply those principles in ethically relevant situations.

Ethical reflection on HCI issues could be improved further by emulating the concept of precedent cases from positive law. This suggestion is inspired by Peterson [10]. Often, different ethical cases have similarities, dealing with structurally similar ethical problems. In HCI research, the infringement of the user's privacy is an issue that applies to many innovations in the field. The same goes for the question of user autonomy [12].

As a scientific community, we can try to initiate a public discussion about the ethical issues of the current research paradigms in HCI. Even though the ethical discussion will not be normatively binding, it might provide researchers in HCI with the kind of normative guidance required to develop technologies that contribute to our society's wellbeing.

Discussion and Future Work

We are well aware that some of the ideas presented in this work have been painted with a broad brush. Thus, a plethora of technicalities will need to be examined in the future.

This is especially true for the relationship between the CODE's general moral imperatives and the middle-range principles. For example, how can the cost-benefit-principle be applied as a method that helps us figuring out how new technologies can be developed in a way that prevents others from being harmed?

In addition, we need to develop a more precise account of how an open debate about the ethical issues in HCI research might work. To give some examples: when are we justified to believe that a consensus on an ethically relevant issue has been reached? Who should participate in the discourse? All worries aside, we believe that a shift of focus towards a more case-study based reasoning is the right way to progress in the ethical debate within HCI research.

Acknowledgements

We gratefully acknowledge the grant 16SV7604 within the project KoBeLU (Context-Aware Learning Environments for Education) from the German Federal Ministry of Education and Research (BMBF).

References

1. Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.
2. Bo Brinkman, Dan Gotterbarn, et al. 2014: Making a Positive Impact: Updating the ACM's Code of Ethics. *Communications of the ACM*, Vol. 59 No. 12, 7-13. <https://doi.org/10.1145/3015149>
3. Bo Brinkman, Don Gotterbarn, Keith W. Miller, and Marty J. Wolf. 2016. All Hands on Deck for ACM Ethics: Updating the Code, Revising Enforcement, Promoting Integrity. *SIGCAS Comput. Soc.* 46, 3: 5–8. <https://doi.org/10.1145/3024949.3024950>
4. Diego Fernando Gutierrez Aponte and Deborah Richards. 2013. Managing Cyber-bullying in Online Educational Virtual Worlds. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death (IE '13)*, 18: 1–18:9. <https://doi.org/10.1145/2513002.2513006>
5. Harry Brignull. 2011. Dark Patterns: Deception vs Honesty in UI Design. <http://alistapart.com/article/dark-patterns-deception-vs.-honesty-in-ui-design>. Last retrieved on Jan. 24, 2017.
6. Oliver Korn. 2014. Context-aware assistive systems for augmented work: a framework using gamification and projection. University of Stuttgart, <http://elib.uni-stuttgart.de/opus/volltexte/2014/9307/>
7. Wallach, Wendell; Allen, Colin. 2009. *Moral Machines. Teaching Robots Right from Wrong*. Oxford University Press
8. Albert R. Jonsen, Stephen Toulmin. 1988. *The Abuse of Casuistry: A History of Moral Reasoning*. University of California Press.
9. Oliver Korn and Alan Dix. 2016. Educational Playgrounds: How Context-aware Systems Enable Playful Coached Learning. *interactions* 24, 1: 54–57.
10. Martin Peterson. 2017. *The Ethics of Technology: A Geometric Analysis of Five Moral Principles*. Oxford University Press.
11. Tom Beauchamp, James F. Childress. 2013. *Principles of Biomedical Ethics*. 7th Edition. Oxford University Press
12. Hauke Behrendt, Markus Funk, and Oliver Korn. 2015. Ethical Implications Regarding Assistive Technology at Workplaces. In *Collective Agency and Cooperation in Natural and Artificial Systems*, Catrin Misselhorn (ed.). Springer International Publishing, 109–130. https://doi.org/10.1007/978-3-319-15515-9_6