Ethical Risks, Concerns, and Practices of Affective Computing

Ethical Risks, Concerns, and Practices of Affective Computing: A Thematic Analysis

1st Deniz Iren
Department of Information Science
Open Universiteit
Heerlen, Netherlands
deniz.iren@ou.nl

2nd Ediz Yildirim
Department of Information Science
Open Universiteit
Heerlen, Netherlands
ediz.yildirim@ou.nl

3rd Krist Shingjergji
Technology Enhanced Learning and Innovation
Open Universiteit
Heerlen, Netherlands
krist.shingjergji@ou.nl

***Abstract*—The recent advances in artificial intelligence (AI) have drawn the attention of the public, policymakers, practitioners, and scientists to the ethical implications of AI. Affective computing is among the sensitive topics, as it deals with human emotions and affect. Research and applications in this field are perceived to raise substantial risks. In this study, we conducted a thematic analysis of the ethical impact statements of 70 papers that are accepted to be presented at the ACII conference. Our aim was to explore how the affective computing research community perceives risks and concerns related to ethics in this field, and how they attempt to address and mitigate these risks. We report our findings of this thematic analysis along with an evaluation of the potential impact of the regulations such as The EU AI Act on the field of affective computing.**

***Index Terms*—affective computing, ethics, thematic analysis**

I. Introduction

In the absence of ethical safeguards, artificial intelligence (AI) runs the risk of perpetuating existing biases and discrimination present in society, exacerbating divisions, and posing a threat to fundamental human rights and freedoms [1]. The rapid advancements in AI have generated significant interest and concern regarding the ethical implications of AI among the general public, policymakers, practitioners, and scientists. A particular area of sensitivity is affective computing, which focuses on understanding and responding to human emotions and affect. Both research and applications in this field are perceived to pose substantial risks [2]. The research community in affective computing emphasizes the significance of ethics by urging authors to incorporate ethical impact statements in their papers. Additionally, they provide guidelines to assist researchers in conscientiously considering the ethical implications of their studies [3].

This study aims to investigate the ethical considerations within the affective computing research community by conducting a thematic analysis of the ethical impact statements of 70 accepted papers for presentation at the Affective Computing and Intelligent Interaction (ACII) conference in 2023. Our primary objective is to gain insight into how researchers perceive and address risks and concerns related to ethics in this field. Furthermore, we discuss the potential implications of regulations such as the EU AI Act [2] on the field of affective computing, providing valuable insights into the regulatory impact on this emerging area.

This research contributes to the ongoing discourse on the ethical implications of AI by offering a thematic analysis of the ethical considerations within affective computing. The findings and evaluation presented in this study serve to inform policymakers, practitioners, and researchers involved in affective computing, facilitating a more nuanced understanding of the ethical landscape and potential regulatory measures.

In this study, we pose the following research questions:

- *RQ1: What are the ethical risks and concerns reported by affective computing researchers?*
- *RQ2: What are approaches proposed by affective computing researchers to mitigate these risks?*
- *RQ3: What is the potential impact of the regulations (e.g., The AI Act) on different types and applications of affective computing?*

This paper is organized as follows. Section II provides background information on various types of affective computing research and practice, and regulations drafted by authorized bodies. Section III describes our thematic analysis research method. Section IV shares our findings. Finally, Section V provides a discussion and concludes the paper.

II. Background

A. Affective Computing Research and Practice Categories

Affective computing encompasses a broad range of technologies and applications spanning various fields. Affective computing systems exhibit diversity in their supported interaction modalities and communication channels. These distinctions hold significance when assessing the ethical considerations associated with affective computing systems, thus, providing a meaningful frame of analysis for our study. In this subsection, we introduce two taxonomies that categorize affective computing systems based on the supported interaction modalities and communication channels.

Affective computing systems comprise unimodal or multimodal interaction modalities. These interaction modalities include *text*, *audio*, *visual*, and *physiological* cues [4] as shown in Fig. 1. Depending on the supported modalities, affective computing systems raise different ethical concerns. For instance, visual data such as facial expressions tend to have more personally identifiable characteristics than text data.

979-8-3503-2745-8/23/$31.00

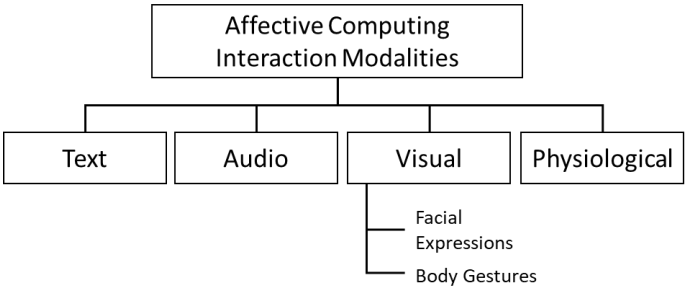

Fig. 1: Simplified taxonomy of affective computing based on supported modalities

Affective computing systems can also be categorized into groups depending on the supported communication channels. This categorization is useful to characterize the application and purpose of the system and proves valuable in examining the ethical risks raised in different cases. The taxonomy of affective computing based on communication channels is depicted in Fig. 2 and elaborated as follows.

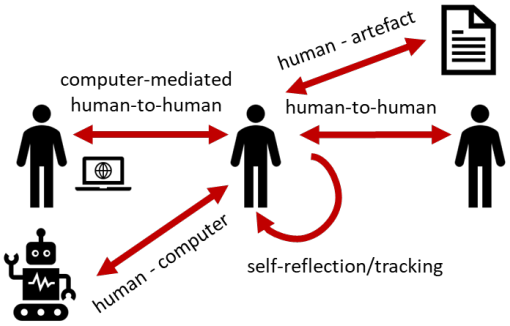

Fig. 2: Taxonomy of communication types

Human-to-human denotes humans communicating in a physical environment supported by affective technologies. *Self-reflection/tracking* refers to perceiving and improving one's own emotional states. *Human-artefact* indicates the analysis of artifacts (e.g., documents) for revealing extra information by considering emotional cues expressed in the artifact. *Human-computer* covers computational systems that can perceive human emotions. Finally, *computer-mediated human-to-human* addresses human communication that takes place on a digital medium such as video conferencing.

B. The AI Act

The AI Act is proposed by the EU to establish a unified regulatory and legal framework for artificial intelligence. This proposal was introduced by the European Commission in April 2021 and received the latest amendment in May 2023. The regulation provides a scheme for a risk-based approach to categorize AI practices as *unacceptable-risk*, *high-risk*, and *low-risk*. Practices that fall under the unacceptable-risk category are prohibited. Examples include manipulating persons through subliminal techniques, exploiting vulnerabilities of special groups such as children and people with disabilities, AI-based social scoring, and remote biometric identification with the purpose of law enforcement. High-risk applications cover systems and practices that pose a risk of harm to health and safety or have potential implications for the fundamental rights of people. They are permitted while being subject to compliance with regulatory requirements and conformity assessment.

The AI Act has several implications for affective computing research and practice. First, it makes a definition of emotion recognition systems to clarify the scope of the term: *"Emotion recognition system means an AI system for the purpose of identifying or inferring emotions, thoughts, states of mind or intentions of individuals or groups on the basis of their biometric and biometric-based data."*

Second, it highlights the reasons for concerns regarding emotion recognition: (a) Emotion expressions and perceptions vary across cultures and contexts, and (b) emotion categories are not reliably associated with a common set of physical/physiological cues. For these reasons, emotion recognition systems run a major risk for abuse and are therefore prohibited to be used in certain situations such as law enforcement, border control, workplace, and education institutions.

Finally, the AI Act defines transparency obligations: systems that recognize emotions based on biometric data must clearly inform their users about it.

III. Methodology

In this study, we used the thematic analysis method to identify patterns of themes within the data [5]. We collected the ethical impact statements of 70 papers that are accepted for publication at the ACII 2023 conference. Subsequently, we coded the data using Atlas.ti, to identify all mentions of *limitations*, *risks/concerns*, and *mitigation* strategies. We grouped the codes into three main categories based on the aspect of the research they are associated with; *study*, *data*, and *application*. Next, we defined the themes inductively by combining the codes depending on their similarity. At this step, we also used the ACII ethical statement guidelines [3] and the concepts addressed by the AI Act [6]. Finally, we categorized each paper using the two taxonomies introduced in Section II-A.

IV. Results

A. Types of Affective Computing

Our analysis covers 70 papers in total. The modalities supported by the affective computing studies or systems reported in these papers are mostly distributed uniformly with the exception of body gestures that were only reported five times (See Fig. 3). 14 papers reported multimodal interaction covering two or more modalities and the rest indicated either unimodal studies or no modality.

A majority of the papers reported studies that support the human-computer communication channel (N=27). 14 papers indicated computer-mediated human-to-human communication. Self-tracking/reflection and human-to-human communi-

Fig. 3: Number of papers grouped by the modality of affective computing

cation were reported seven times each. Finally, human-artefact type was indicated only two times (See Fig. 4).

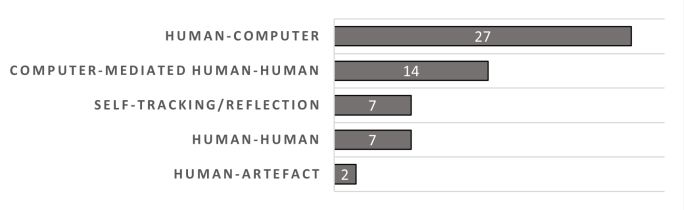

Fig. 4: Number of papers grouped by the communication channel addressed in the study

B. Thematic Analysis Results

We identified 40 unique codes that represent limitations, risks, and concerns, and 42 that indicate mitigation strategies and good practices. These are grouped under three categories (i.e., study, data, and application) and labeled with seven themes (i.e., human subjects, research design, environmental impact, data quality, nature of data, data accessibility, and application. The outcome of the thematic analysis is shown in Table I. The numbers in parentheses depict the number of papers a particular code was assigned to, therefore signifying its importance or frequency.

V. Discussion and Conclusion

The impact of the AI Act: Potentially, the AI Act has a critical impact on affective computing research and practice. Even though currently, scientific research is not subject to regulation, the outcomes of affective computing research will be. Systems that use biometric-based data to infer emotions are considered high-risk [6]. This means most studies that use modalities mentioned in Fig. 3 other than text will be subject to regulation and audit. Furthermore, affective computing systems that operate in critical domains (e.g., healthcare (20), education (4), and social services (9)) will be prohibited. Due to the transparency requirements mandated by the AI Act, all systems that recognize emotions will need to inform their users in full transparency. The AI Act highlights the reasons why emotion detection raises serious concerns; emotional displays are context and culture-dependent, and physiological cues do not reliably match with inferred emotions. As the affective computing research community, we must prioritize studies that address these shortcomings to alleviate the concerns.

Ethical risks and mitigation strategies: Our findings indicate that several risks are frequently acknowledged and mitigation strategies are commonly relied upon. When a study involves ***human subjects***, most researchers seek Institutional Review Board (IRB) approval and apply good practices such as informed consent and allowing participants to abandon the study at will. Regarding ***study design***, researchers acknowledge limitations such as *context-specificity* and propose to improve the study as future work. Only a few studies report the ***environmental impact***. Thus, there is room for improvement in raising awareness regarding the environmental impact of research.

Data quality is one of the most reported aspects in the ethical impact statements. Researchers often acknowledge size and diversity limitations and suggest using larger and more diverse data in future studies. There is a strong awareness regarding biases in data that originate from various sources [7]. Identifying and handling biases in data has become a regular practice. The ***nature of data*** can be sensitive, personal, and private, thus, requiring safekeeping, anonymization, and de-identification. Several researchers suggest that emotion data must be considered sensitive and handled with the same care as healthcare data. Publishing ***open data*** contributes to scientific reproducibility. Published datasets must have an explicitly defined license and a user agreement. Our findings suggest that there is a common misconception that the use of public datasets ensures generalizability and overcomes privacy concerns. When public datasets are used, the researcher who uses the dataset is responsible for checking the ethical aspects regarding the original study that created the dataset. Thus, the used dataset's characteristics should be reported. If such information is not available in the original publication of the dataset, the researchers need to explore the data for biases, balance, and population representability.

Finally, there are important concerns regarding the misuse of research outcomes and harmful ***applications***. Keeping users informed regarding the affective computing system and addressing the failure scenarios of the application are reported as two ways to tackle these concerns. Furthermore, several researchers call for governmental regulations.

Ethical impact analysis guidance: Despite the substantial steps taken by the affective computing community toward guiding researchers to critically think and openly report the ethical impact of their studies, we observe that the ethical impact statements are sometimes incomplete and reported in a non-standard manner. More detailed guidelines are required for examining and reporting the ethical impact of affective computing research.

Limitations: This study is not without limitations. We examined only the ethical impact statements and abstracts. Therefore, our observations reflect what is reported by the authors in the ethical impact statement rather than their actual research practices reported in other parts of the papers. Additionally, this work analyzes the accepted papers of ACII 2023. Even though this is a good representation of the affective computing community, still, the scope is limited to one venue

TABLE I: THEMATIC ANALYSIS OUTCOME

THEMES		CODES		
		LIMITATIONS	**RISKS**	**MITIGATION**
STUDY	**HUMAN SUBJECTS**	➪ Participant selection and compensation (3)	➪ Limited oversight (2) ➪ Harm to participants (2)	➪ Involve IRB(26) ➪ Apply informed consent (22) ➪ Participants can drop-out at will (4) ➪ Transparent reporting (2)
	STUDY DESIGN	➪ Context-specific (2)	➪ Results are not generalizable (6) ➪ Reduced construct validity (2)	➪ Improve the study (5) → Conduct more research (4) → Improve the performance (3)
	ENVIRONMENTAL IMPACT		➪ Environmental Impact (5)	➪ Examine and report environmental impact (2) ➪ Train small models (1) ➪ Use pretrained models (1) ➪ Avoid over-personalization of models (1)
DATA	**DATA QUALITY**	➪ Small sample size (10) ➪ Sample is not representative (4) → Demographics (4) → Limited set of emotions (1) ➪ Data imbalance (2)	➪ Results are not generalizable (6) ➪ Discrimination (3) ➪ Biases (24) ➪ Reduced accuracy (3)	➪ Improve the data (10) → Collect more data (7) → Collect more diverse data (4) → Apply sampling strategies (2) → Balance data (3) → Examine the biases (4) → Use multiple datasets (2)
	NATURE OF DATA		➪ Sensitive data (5) → Healthcare/mental → Offensive content ➪ Private data (14) ➪ Personally identifiable data (1) ➪ Unauthorized access to the data (2) ➪ Unclear IP rights and licensing (2)	➪ Anonymization/De-identification (22) ➪ Setup data protection policy (2) ➪ Establish data protection measures (2)
	OPEN DATA	➪ Private/unavailable research data (2)	➪ Reproducibility is hindered ➪ Misuse of data	➪ Make research data available (5) ➪ License the published datasets (2) ➪ Establish EULA for published datasets (2)
APPLICATION	**APPLICATION**	➪ Limited stakeholder involvement (2) ➪ Critical domains and application fields → Healthcare (20) → Education (4) → Social services (9) → Law enforcement and border control (0) → Workplace (2)	➪ Harmful applications (18) → Surveillance → Deception → Manipulation → Restrict autonomy ➪ Societal adverse impact (2) → Limit fundamental rights → Controversial subjects ➪ Failure consequences (1)	➪ Identify and address failure consequences (1) ➪ Provide transparent information to user (2)

and it needs to be extended to cover other venues.

Future work: We think that the analysis of ethical risks and mitigation strategies is important and timely for our community. We will extend our work to cover IEEE Transactions of Affective Computing. Also, we plan to call for collaborators in our endeavor to create a systematic set of guidelines for evaluating and mitigating the ethical impact within affective computing research. Finally, we plan to communicate our findings as well as our future discussions at the conference to the EU committee that is responsible for creating the AI Act with the purpose of opening a dialogue channel between our community and policymakers.

Ethical Impact Statement

The analyzed papers could not be cited in our work at this point due to the double-blind review policy. We will attempt to provide the necessary citations in later versions if possible. To the best of our knowledge, this thematic study of the literature has no other substantial ethical impact.

References

[1] *Unesco's recommendation on the ethics of artificial intelligence: Key facts*, https://unesdoc.unesco.org/ark:/48223/pf0000385082.locale=en, accessed 18-June-2023, 2021.

[2] *Artificial intelligence act*, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206, accessed 18-June-2023, 2021.

[3] D. Ong, J. Hernandez, R. Picard, *et al.*, *Writing an ethical impact statement for acii2023*, https://acii-conf.net/2023/wp-content/uploads/2023/03/instructions-ethical-statement.pdf, accessed 18-June-2023, 2023.

[4] Y. Wang, W. Song, W. Tao, *et al.*, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Information Fusion*, 2022.

[5] V. Clarke, V. Braun, and N. Hayfield, "Thematic analysis," *Qualitative psychology: A practical guide to research methods*, vol. 3, pp. 222–248, 2015.

[6] *Artificial intelligence act - dract compromise ammendments*, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf, accessed 18-June-2023, 2023.

[7] B. Aysolmaz, D. Iren, and N. Dau, "Preventing algorithmic bias in the development of algorithmic decision-making systems: A delphi study," 2020.