Guest Editorial: Ethics in Affective Computing

STUNNING advances in machine learning are heralding a new era in sensing, interpreting, simulating and stimulating human emotion. In the human sciences, research is increasingly highlighting the explanatory power of emotions, feelings, and other affective processes to predict how we think and behave. This is beginning to translate into an explosion of applications that can improve human wellbeing including methods to reduce stress and improve emotion regulation skills, techniques to support healthier social media use, pain monitoring in neonates, and decision-support tools that recognize emotional bias.

Yet these transformations raise legitimate concerns. Affective computing applications sometimes proceed independently from findings in affective science or a broader consideration of human well-being. For example, some "emotion recognition" products claim to reveal what a customer truly feels from a decontextualized image of their face alone (a capability that affective science has long deemed implausible). Other methods naively incorporate biases that would undermine individuals' rights or fair access to resources. Even when a product has a strong scientific basis, there can be good reasons to restrict its use due to questionable societal value or concerns that its widespread use could undermine cherished rights.

These concerns are receiving intense attention from the media and government. Yet calls for regulation can also proceed independently from findings in affective science or a broader consideration of human well-being. For example, the European Parliament recently adopted a negotiation position on the Artificial Intelligence Act that bans the use of emotion recognition systems in the workplace, and in educational institutions. But the rationale for such bans sometimes confuses the underlying technology. Thus, regulators sometimes equate emotion recognition with face recognition, whereas only the latter necessarily reveals an individual's identity. Concerns about recognizing emotion from the face alone are generalized to any affect recognition application, even to evidenced-based methods that incorporate multiple modalities and context when making inferences.

Such controversies highlight a greater need for dialogue between those creating affective computing systems and those concerned about the ethical and legal impacts. As a field, it is incumbent on us to build awareness of what the technology does, its limitations, and how it engages with human values. To this end, we called for articles from those within and outside the affective computing community on how we build ethical technologies by considering benefits, risks, and use cases. In response to an open call, we received eighteen submissions from which we selected the following seven articles.

Featured Articles: The articles cover a range of ethical concerns. Some authors address fundamental issues that would impact any application, while others deeply consider the issues unique to a specific domain, like education. Finally, authors consider a variety of perspectives to avoid abuse and promote well-being.

Privacy is a central concern in affective computing and the focus of several of our articles. People freely express their emotions to facilitate communication and coordination with others (even machines), yet they often regulate these expressions to hide their true feelings (e.g., to avoid hurting others' feelings or to protect against potential exploitation). To the extent that machines could "see" through these regulatory attempts, they might have a superhuman ability to predict an individual's goals, values or action tendencies. The first four articles struggle with these privacy concerns.

Aijaz, Kumar, Chattar, Shukla, and Mutharaju begin the special section by arguing that affective computing is shifting the balance of power between humans and machines. In the past, machines were designed to be transparent but humans were largely opaque to machines. But as deep learning methods become increasingly opaque (even to their designers) and machines gain the ability to recognize "true" emotions, this dynamic may be shifting, potentially leading to adverse repercussions, such as infringement of autonomy, deception, and manipulation.

The second article by Branford shares this concern and argues that affective computing necessarily requires access to private information in order to achieve its aims. The author highlights a long history of scholarship in emotion research that context matters, and then raises the question as to whether incorporating context necessitates intrusive surveillance. Building on research that emphasizes the importance of context in emotion recognition, Branford argues that this contextual information will necessarily involve sensitive information about a person's activities and relationships and calls for a "situated ethics" to critically examine these concerns.

Whereas the first two articles emphasize privacy concerns, the next proposes concrete solutions. Uddin, Canavan and Yin consider that there are two data sources for affect sensing; the primary raw input data (e.g., video) and the secondary outputs of basic machine learning models (e.g., facial points or predicted facial muscle actions). There are obvious ethical benefits for using secondary data, as it does not contain readily identifiable face and voice data and can thus be used for de-identified affective computing. The authors use a spatial-temporal graph

Date of current version 1 March 2024.
Digital Object Identifier 10.1109/TAFFC.2023.3322918

representation of the secondary outputs, and present how aspects of graph network sciences can be used to interpret the affective data, for example by considering the centrality of a node in the graph which represents a particular facial muscle.

Behnke, Saganowski, Kunc, and Kacienko propose concrete methods to address the privacy concerns around the explosive growth in wearable technology. Wearables create the opportunity to collect and study data related to emotion in daily life at a scale never before possible, yet ethical conversations have not kept pace with this growth. The authors identify specific risks of wearable technology together with recommendations of strategies to minimize risks. Building on interviews with numerous experts from across the ethics and research community, they propose a checklist for designers to guide application development.

Hupont, Tolan, Frau, Porcaro, and Gómez focus on bias and how this can arise from a lack of diversity within affective computing. The authors use bibliometric methods to examine diversity across 10 years of research in leading AC publication venues. In documenting their insights, the authors observe low, but slowly increasing levels of representation of women in AC research and a lack of geographic representation from Latin American and Africa too. Driven by these insights, the authors formulate detailed strategies for increasing diversity in AC research.

The final two articles explore how unique ethical concerns arise in specific domains. Melhart, Togelius, Mikkelsen, Holmgård and Yannakakis focus on computer games, which are estimated to be played by about 40% of the world's population. Gaming companies are investing considerable effort to understand and influence players' emotions to enhance player enjoyment but also to increase profits. Yet there has been relatively little written about gaming, AI, and ethics. This paper surveys the current state of the art with regards to AI in games, discusses ethical considerations of these systems using the affective loop: elicitation, sensing, detection, and adaptation, and proposes concrete mitigations for the ethical risks.

Finally, Banzon, Beever, and Taub examine educational applications, one of the areas potential EU regulations seek to ban. Recent advances in affective computing have made it possible to detect student boredom or frustration to guide targeted instruction, but less consideration has gone into addressing the ethical concerns related to detecting and reporting on learners' emotions specifically within applied educational contexts. The authors discuss the use of Reflexive Principlism as a way to help guide engineers to think through ethical issues. Inspired by core ethical principles such as benevolence and nonmaleficence, Reflexive Principlism provides a set of concrete decision-making processes to help engineers work through how these principles arise in a specific application.

Conclusions: These seven articles forge new pathways forward, not only in their findings, but also in setting examples for how affective computing researchers, and others deploying this technology, can proceed with ethical practice. In pulling this collection together, we re-engage with the original goals of this field to create computing that respects the whole human being – not reducing any person to just the actions of a "user," but restoring the rightful order of innovating technology in service of human needs, honoring people with their complete minds and bodies, hearts and souls.

JONATHAN GRATCH
University of Southern California
Los Angeles, CA 90007, USA
gratch@ict.usc.edu

GRETCHEN GREENE
Hastings Center, Garrison
NY 10524-4125, USA
kgretchengreen@gmail.com

ROSALIND PICARD
MIT, Cambridge
MA 02139, USA
picard@media.mit.edu

LACHLAN URQUHART
University of Edinburgh
EH8 9YL Edinburgh, U.K.
lachlan.urquhart@ed.ac.uk

MICHEL VALSTAR
University of Nottingham
NG7 2RD Nottingham, U.K.
Michel.Valstar@nottingham.ac.uk

ACKNOWLEDGMENTS

We express our sincere thanks to all the authors and reviewers for extending their cooperation in revising and in preparing the final versions of the papers.

Jonathan Gratch is a research full professor of computer science and psychology with the University of Southern California (USC) and director with Virtual Human Research, USC's Institute for Creative Technologies. His research focuses on computational models of human cognitive and social processes, especially emotion, and explores these models' potential to advance psychological theory and shape human-machine interaction. He is the founding editor-in-chief (retired) of I*EEE's Transactions on Affective Computing*, associate editor of Affective Science and the *Journal of Autonomous Agents and Multiagent Systems,* and former president with the association for the Advancement of Affective Computing (AAAC). He is a fellow of AAAI, AAAC, and the Cognitive Science Society.

Gretchen Greene received the CPhil, MS, and JD degrees. He is an international AI policy advisor, lawyer, and computer vision scientist, developing AI, privacy, and human rights policy guidance and industry best practices. A former U.S. national lab mathematician and law firm partner, Greene has worked for the U.S. Departments of Energy, Defense, Justice, and Homeland Security, the U.K. House of Lords, MIT, Harvard, and Partnership on AI. Greene has been interviewed on AI policy and face recognition by the BBC and Politico Europe, testified on biometrics bills, written AI guidance read in ∼8000 cities, developed algorithms for autonomous car navigation and terrorist tracking, and published in machine learning, policy, and science journals. Greene is a senior advisor at The Hastings Center, and CEO at Greene Strategy.

Rosalind Picard (Fellow, IEEE) received the BS degree in electrical engineering from the Georgia Institute of Techology, and the SM and ScD degrees in electrical engineering and computer science from MIT. She wrote the book Affective Computing in 1997 and has authored hundreds of peer-reviewed science and engineering articles. She is a professor with the MIT Media Lab, where she directs Affective Computing Research. She co-founded two startups Affectiva and Empatica. She is a member of the National Academy of Engineering and the National Academy of Inventors, and a fellow of the ACM, APA, AAAC. She received the 2022 International Lombardy Prize for Computer Science Research, called the "Nobel" in Italy.

Lachlan Urquhart is a senior lecturer with Technology Law and Human-Computer Interaction, University of Edinburgh. He is founder of the Regulation and Design Lab, and a director with the Scottish Research Centre for Intellectual Property and Technology Law and the Centre for Research on Information, Surveillance, and Privacy. He has published 60 papers and has been awarded more than £6m in externally funded research projects. His research explores the socio-technical, legal, and ethical aspects of designing, living with, and regulating emerging technologies e.g., smart homes, Internet of Things, artificial intelligence.

Michel Valstar received the master's degree in electrical engineering from the Delft University of Technology, in 2005, and the PhD degree in computer science with Imperial College London, in 2008. He is founding CEO of BlueSkeye AI and an honorary professor with Automatic Human Behaviour Understanding in computer science with the University of Nottingham. He is an expert in automatic facial expression analysis. Over 18 years in this field his work has been cited more than 16000 times in more than 100 peer reviewed publications. He is the organiser of 9 international conferences on the subject. His work has received extensive media coverage including in; Science Magazine, The Guardian, New Scientist and on BBC Radio and television.