



Why we need to be weary of emotional AI

Mantello Peter¹ · Manh-Tung Ho^{1,2,3}

Received: 21 August 2022 / Accepted: 26 September 2022

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Artificial intelligence (AI) is rapidly transforming how we work and live. One of the fastest growing areas of its development belongs to intelligent machines that can sense, read and evaluate human emotion. More commonly known by its commercial moniker, emotional AI, the technology is quickly becoming an integral layer in smart city design. Its origin traces back to the groundbreaking work of Rosalind Picard in affective computing over twenty years ago. Picard coined the term ‘affective computing’ to describe her development of machine intelligence that can respond to a person’s psycho-physical state. Although in its early commercial stage, emotional AI is already a lucrative, 24 billion dollar industry whose profits are expected to double by 2024 (Crawford 2021).

Evolving in ever greater sophistication and complexity, emotion-sensing devices are now featured in autonomous cars, classroom teaching aids, smart toys, home assistants, online conferencing, email software, advertising kiosks and billboards, fast food and drive-through menus, care robots as well as public and private security systems. Unlike other AI applications that rely on extracting data from a person’s corporeal exterior, emotional AI passes into the interior and highly subjective domain of a person via biometric means. This includes the use of algorithms, biosensors, and actuators that harvest non-conscious data gleaned from someone’s heartbeat, respiration rate, blood pressure, voice tone, word choice, body temperature, galvanic skin responses, head, and eye movement, and gait. More advanced affect tools incorporate state-of-the-art machine learning, big data and

natural word processing to allow for greater degrees of accuracy, flexibility, and personalization as well as situational and temporal context.

Like most AI technologies, affect recognition devices promise to augment and enhance daily existence. But as a far more invasive genus of surveillance capitalism, the technological adoption of emotional AI is problematized by a myriad of legal, ethical, cultural, and scientific issues. In this short essay, we identify five major tensions in the infusion of emotional AI in society.

First, is the technology’s reliance on stealth data tracking, which may lead to unethical or malicious misuse. Affect tools are designed to harvest intimate data from an individual’s subjective state without necessarily their awareness or permission. This creates multiple possibilities for its malicious or harmful misuse. For example, emotion-sensing devices in the workplace may lead to bias or discrimination against a worker for their lack of ‘attitudinal conformity’. In turn, affect-sensing tools may lead to emotional policing, coercing a worker to always be happy, authentic, and positive. At the same time, they diminish their ability to backstage their feelings as well as foment higher levels of anxiety, stress, and resentment. Similarly, affect tools in automobiles may lead to unfairly higher car or health insurance premiums. Concomitantly, in commercial settings, individuals may be exposed to empathic surveillance without their knowledge and depending on country, consent. For instance, AdMobilize links their AI-driven software to public transit security cameras which then monitor audience responses to interactive ads. Besides analyzing gender, age, and dwell time, the software uses facial analysis to detect micro-expressions of surprise, happiness, discontent, and neutrality. The goal of AdMobilize’s affect tools is to assess ad performance and customer engagement.

Second, are cultural tensions arising from these emotional AI technologies crossing national and cultural borders. Although emotion-sensing technologies are predominantly designed in the West, they are being sold to a global marketplace. Problematically, as these devices cross international borders, their algorithms are seldom tweaked

✉ Manh-Tung Ho
tung.homanh@phenikaa-uni.edu.vn

¹ Ritsumeikan Asia Pacific University, Beppu, Oita 874-8577, Japan

² Centre for Interdisciplinary Social Research, Phenikaa University, Yen Nghia Ward, Ha Dong District, Hanoi 100000, Vietnam

³ Institute of Philosophy, Vietnam Academy of Social Sciences, 59 Lang Ha Street, Ba Dinh District, Hanoi 100000, Vietnam

for racial, cultural, ethnic, or gender differences. A growing body of research shows that AI models that do not allow for difference or diversity can lead to unintentional bias or false positive identification, negatively impacting a target individual. This problem is further compounded by the lack of international consensus on the values and ethics that should be encoded into intelligent machines as well as cross-cultural incongruences arising from a country's legal understanding of privacy. For instance, while facial recognition and social credit systems are banned in many Western countries, China faces far less push-back because the notion of collective security is valued more than individual privacy (Mantello et al. 2021). Additionally, Chinese citizens are found to show greater trust in government-sponsored data collection than their Western counterparts (Roberts et al. 2021).

Third, is the lack of industry standard. Like the hidden data-gathering activities of many smart technologies, emotional AI will be far harder to collectively regulate as it is being developed as a proprietary layer in many products. A prime example is the automotive industry. Companies, such as Ford, Porsche, Audi, Hyundai, Toyota, Honda, BMW, Volkswagen, and Jaguar, in the name of safety and comfort enhancement, are developing in-cabin concierge systems that can track and respond to the emotional states of drivers (McStay and Urquhart 2022). Yet as researchers McStay and Urquhart (2022) observe, for the automobile industry, algorithmic secrecy is imperative for maintaining a competitive edge. This means that algorithmic transparency and collective standards for non-conscious biometric data collection will not occur for some time.

Fourth, existing ethical frameworks for emotional AI are often vague and inflexible. This is due to various businesses in different cultural settings having differing rationales or goals for adoption of the new technology. For example, the Japanese voice analytics company, Empath, sees the technology as a way for call centers to optimize workplace productivity by providing supervisors with a panoptic window into the subjective state of each member of their customer service team. On the other hand, Moodbeam's emotion bracelet offers companies a neoliberal alternative to the far more administrative and costly worker wellness programs. As the company's promotional literature suggests, wearing the affect-sensing bracelet will enable workers to automatically share data of their subjective state with both managers and co-workers. This neoliberal approach to mindfulness alludes to the premise that 'sharing is caring.' Besides varying objectives for adoption comes the practical limitations of implementation and establishment of concrete metrics for measuring the technology's effectiveness. But ensuring the efficacy of emotional AI technology requires having full-time staff skilled in data analytics and data management. Yet many companies are implementing emotion-recognition

systems without personnel skilled in data analytics and data management experts.

Last, but not least, comes the shaky science of the emotion-recognition industry (Barrett 2017; Crawford 2021). A growing number of critics argue how emotions can be made computable when the science community cannot agree on exactly what emotions are, how they are formed or how they manifest themselves. Are emotions hard-wired into the psycho-physical makeup of an individual or socially and culturally contingent? Added to this debate is the fact that leading emotional AI companies still rely on Paul Eckman's the now discredited 'universality of emotions' (Barrett 2017). Pushing back against these arguments are the engineers who insist emotions are computable, and that any limitations in diversity or cultural affordance will ultimately be solved by better algorithms.

As emotional AI becomes more pervasive in society, it will have profound impacts on the daily lives of citizens. Technology that endeavors to make transparent the inner recesses of a person's being raises critical questions about data privacy in public spaces, empathic monitoring and control as well as how regulatory mechanisms should best ensure the best interests of society. Thus, this essay provides an overarching framework for discussion of the social, legal, and ethical implications of emotional AI technology. The five major tensions highlighted in this article must be thoroughly addressed in order for individuals to live well and ethically in this new era of human-machine relations.

Acknowledgements This study is part of the project "Emotional AI in Cities: Cross Cultural Lessons from UK and Japan on Designing for an Ethical Life" funded by JST-UKRI Joint Call on Artificial Intelligence and Society (2019), Grant No. JPMJRX19H6. The authors would like to thank all the APU faculty members that helped us distribute the survey. Author Manh-Tung Ho would like to express his sincere gratitude for the support of the SGH Scholarship Foundation.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Data availability statement There are no data associated with this paper

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

- Barrett LF (2017) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc Cognit Affect Neurosci* 12(1):1–23. <https://doi.org/10.1093/scan/nsw154>
- Crawford K (2021) Artificial intelligence is misreading human emotion. <https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/>. Accessed 2021 May 14
- Mantello P, Ho M-T, Nguyen M-H, Vuong Q-H (2021) Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward emotional AI in the workplace. *AI Soc*. <https://doi.org/10.1007/s00146-021-01290-1>
- McStay A, Urquhart L (2022) In cars (are we really safest of all?): interior sensing and emotional opacity. *Int Rev Law Comput Technol*. <https://doi.org/10.1080/13600869.2021.2009181>
- Roberts H, Cowls J, Morley J, Taddeo M, Wang V, Floridi L (2021) The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI Soc* 36(1):59–77

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.