



Universiteit
Leiden
The Netherlands

Emotional privacy: a philosophical investigation of emotion recognition technology

Prégent, A.P.

Citation

Prégent, A. P. (2026, September 2). *Emotional privacy: a philosophical investigation of emotion recognition technology*. Retrieved from <https://hdl.handle.net/1887/4309114>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4309114>

Note: To cite this publication please use the final published version (if applicable).

7. General conclusion

This dissertation has examined the ethical, social, and normative implications of fully functional ERTs. While current systems remain flawed, contested, and in many cases demonstrably unreliable, the ongoing developments in affective computing, together with sustained commercial, political, and institutional interest in ERTs, suggest that fully functional ERTs may be successfully developed in the near future. In light of this trajectory, the guiding approach of the thesis has been anticipatory. Rather than examining the ethical, social, and normative consequences of existing systems, the dissertation has asked what would follow if ERTs were to work as intended and which social and normative structures would be challenged by technologies capable of reliably detecting affective states.

By assuming the possibility of fully functional ERTs, the thesis has shown that their integration into everyday social life would risk unsettling many of the core expressive, epistemic, and normative conditions on which interpersonal and societal functioning depend. A close examination of existing privacy frameworks reveals that mainstream accounts of informational privacy currently lack the conceptual resources to capture these harms. This exposes a conceptual and normative gap in contemporary privacy theory. The dissertation addresses this gap by introducing and defending emotional privacy as a subconcept within the informational privacy cluster. Emotional privacy protects individuals' expressive ability, and thereby the conditions under which individuals can modulate and negotiate the legibility of their affective life *in situ*.

This protection secures at least three core socio-normative interests that are dependent on individuals' expressive ability. These are (i) self-presentation (individual level), (ii) role plurality and adherence to role-bound norms of expressiveness (interpersonal level), and (iii) social regulation and cohesion (societal level). By articulating these functions, the thesis demonstrates that emotional privacy identifies a class of harms that is distinct from, yet compatible with, standard informational privacy concerns.

On this conceptual foundation, the dissertation develops emotional privacy into a candidate moral right. Through the use of technomoral scenarios and the evaluation of ERTs via Contextual Integrity and ALTAI-style assessments, the thesis shows that recognising a moral right to emotional privacy alters the normative weight of privacy in ways that can reshape or even shift the ethical verdicts reached about affective technologies.

The central claim defended throughout the dissertation is thus the following: when technologies can reliably infer emotional states, the resulting harms are not confined to data privacy, manipulation, or the exploitation of sensitive information. A distinct, but interconnected, set of harms emerges, centred on the expressive, relational, and normative structures on which social life depends. Ethical and regulatory assessments of ERTs must therefore consider not only familiar informational privacy risks but also the risks of undermining the social infrastructure that makes self-presentation, role negotiation, and social cohesion possible.

Main contributions of the dissertation

One central contribution of the dissertation is its conceptual and ethical focus on fully functional ERTs. By shifting attention from current systems to fully functional ERTs, the dissertation moves beyond current discussions, which largely criticise existing ERTs for their unreliability, bias, and lack of scientific grounding. To date, little scholarship has examined what would follow, ethically, if these technical limitations were overcome. Hence, this dissertation is one of the first to question the ethical concerns that would follow from these systems working as intended, providing ethical guidance ahead of technological development.

A second contribution is the development of an integrative and function-first account of emotional life, articulated across Chapters 3 and 4. This account provides the theoretical foundation necessary for understanding what is at stake when technologies target affective states. At the core of this contribution is the proposal of EEs as serving two different kinds of functions for humans and machines. While humans rely on embodied, contextual, and norm-grounded informational signals, ERTs are assumed to rely on probabilistic regularities in sets of physiological and behavioural changes, thereby creating a second, distinct mode of recognition. This establishes that ERTs operate through a fundamentally different epistemic route into affective life. This bi-functionality thus creates a conceptual bridge between work in philosophy of emotion and the ethics of emerging technologies.

A third contribution lies in the introduction and defence of emotional privacy as a subconcept within the informational privacy cluster. Using conceptual engineering, the dissertation argues that emotional privacy protects not only actual control over particular disclosures but also the background expressive ability that enables individuals to modulate what their expressions count as across contexts. This dual-level structure clarifies why the harms posed by fully functional

ERTs arise not only from the loss of actual control over one's emotional expressions but also from the defeat of the expressive ability that underlies the production of these emotional expressions.

A final major contribution of this thesis is the normative development of emotional privacy into a candidate moral right within the informational privacy cluster. By comparing emotional privacy to existing theories of informational privacy, contextual integrity, and mental privacy, the dissertation shows that emotional privacy captures a distinct normative domain: the maintenance of expressive discretion against technologies that make affective life continuously legible. Fully functional ERTs, on the analysis offered here, restructure the affective and expressive fabric of social interaction, rather than merely extending ordinary human mindreading. The recognition of a moral right to emotional privacy thus adds a new normative dimension to the ethical and regulatory evaluation of affective technologies.

Some limitations and conceptual challenges

The dissertation faced some limitations and conceptual challenges. First, the potential impact of this thesis may seem, and is to a certain extent, restricted by its core assumption concerning the prospect of fully functional ERTs. It remains possible that technological development will never reach the level of reliability assumed here, or that progress unfolds along trajectories that diverge from those reconstructed in this work. Nonetheless, these limitations are not fatal to the central argument, since the conceptual and normative findings apply to any technology that produces a comparable increase in affective legibility, irrespective of whether it ultimately matches the functional assumptions described in the preceding chapters. This further implies that while removing the issues related to the technical limitations of current ERTs was necessary to isolate the specific kind of harms at stake, the concept of emotional privacy may also be relevant for the analysis of partially functional ERTs, provided that they increase affective legibility in ways that pressure existing expressive practices.

A second limitation concerns the adoption of an integrative perspective on emotions and emotional expressions. In unifying insights from divergent disciplinary approaches, the dissertation necessarily forgoes deeper engagement with foundational debates about the nature of emotions and emotional expressions, debates for which the “right” answers remain contested across philosophy, psychology, and affective science. The integrative and function-first account

developed here recognises those quandaries but deliberately brackets most of them in order to concentrate on their functional roles. While this choice was necessary to pursue the aims of the dissertation, it leaves unresolved ongoing questions about the nature of emotions and emotional expressions.

A further conceptual challenge arises from the tension between efforts to empirically ground the assumptions about fully functional ERTs and the anticipatory orientation of the project. Since it is still unclear how fully functional ERTs will be developed, I have been careful to stay at a high level of technical explanation regarding the recognition process of fully functional ERTs. Nonetheless, the anticipatory orientation forces the thesis to commit to some views concerning the technical development of technology-based emotion recognition, which might turn out to be erroneous or irrelevant in the near future. I have taken this level of uncertainty into account, and this is why Chapters 4 and 5 do not tie their conclusions to any particular technique regarding the recognition process, only to the idea that the desired outcome is reached, namely that the technology works as intended.

Potential directions for future research

Future research is needed in affective computing and regarding the use of fully functional ERTs. In direct continuation to this thesis, one interesting and needed line of research concerns the legal and political implications of a moral right to emotional privacy. While the dissertation establishes its normative foundations in privacy theory, further work should assess how such a right can be operationalised within existing legal frameworks, particularly in relation to data protection law, human rights instruments, and national and supranational regulations. Questions concerning enforcement mechanisms, emotional privacy as a legal subcategory of the human right to privacy, and the balance between emotional privacy and other collective interests (for example, in public health, education, or security contexts) appear to be primary candidates for further investigation. Addressing these issues would provide more concrete guidance for policymakers and regulators navigating the rapidly evolving landscape of affective technologies.

A second avenue for future research concerns the empirical and socio-technical dimensions of emotional privacy. The thesis has offered a largely Westernised philosophical and conceptual analysis; it would be valuable to investigate how individuals in different social and cultural contexts understand, value, and attempt to secure emotional privacy in practice. Empirical

studies on users', developers', and regulators' perceptions of affective legibility could help refine the concept of emotional privacy, test its intuitiveness, and reveal further mechanisms through which expressive discretion is preserved or undermined in everyday technosocial environments.

A third direction lies in extending the analysis to other affect-sensitive technologies beyond ERTs, such as social robots, conversational agents, and immersive environments that adapt to users' moods. These systems may not always be marketed as "emotion recognition" tools, yet they increasingly rely on affective inferences to guide interaction. Given the rise of affective computing in HCI, exploring how emotional privacy applies to these neighbouring technologies, and whether additional conceptual resources are needed, would deepen our understanding of the broader affective ecosystem in which future interactive technologies are likely to operate.

In closing, this dissertation has argued that as technologies grow ever more capable of rendering affective life transparent, the task of safeguarding the conditions under which we appear to one another as social and moral agents becomes increasingly urgent. Emotional privacy, as developed here, is not merely a barrier against intrusion but a tool to protect the opacity, ambiguity, and selective self-disclosure that are integral to humans and society's well-being. Whether fully functional ERTs ultimately emerge or not, the questions raised in this thesis speak to a broader challenge of our time: how to preserve the spaces in which persons can shape who they are, how they are seen, and how they relate to others within increasingly mediated worlds. Safeguarding those spaces is not only a matter of protecting information from malicious use but of protecting the social fabric in which meaning, agency, and shared life are formed.