



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).

English Summary

In recent years artificial intelligence has moved decisively into the domain of affect. Technological breakthroughs, together with commercial enthusiasm and political interest, have accelerated this expansion. Affective monitoring cameras are now present in public spaces. Emotion sensitive tutoring systems are being introduced in classrooms. Workplaces experiment with analytics that track employees' positivity (or lack thereof). And border control authorities explore systems that allegedly detect deception through micro-expressions. Although many of these systems are scientifically contested and face significant technical limitations, emotion recognition technologies (ERTs) are making impressive leaps forward, and this continuous progress raises an urgent philosophical question: what would happen to social life, to our expressive practices, and to our normative expectations if ERTs were to function as intended and reliably track our affective lives in real time?

This is the question guiding the dissertation. Rather than evaluating current (and flawed) systems, it adopts an anticipatory ethical approach anchored in real sociotechnical developments to map the potential disruptions that would follow from reliable, continuous, and ubiquitous affective legibility. The prospect of fully functional ERTs allows us to anticipate this world and the new kinds of harms that would arise from their development. After investigating the ways in which emotional expressions currently function in social interactions, the dissertation proposes and articulates a novel conception of emotional privacy, which has for purpose to provide an additional (normative) conceptual tool to better grasp the stakes at play in the deployment and use of ERTs for both individuals and societies.

The main body of the thesis is structured into three parts across five chapters. **Chapter 2** develops the first part by reconstructing the technosocial trajectory of ERTs. It follows their evolution from early face-based systems that relied on simplistic emotion models and static images to context-sensitive systems that integrate multimodal inputs to detect individuals' emotional states. In the process, the chapter identifies persistent technical challenges that currently significantly limit these technologies, even as ERTs continue to be deployed across many domains of social life, including retail, education, policing, border control, healthcare, human resources, self-tracking applications, and robotics. The chapter concludes with a discussion of regulatory initiatives, with a particular focus on the EU AI Act and its treatment

of ERTs. While these initiatives explain the pressing need to strictly regulate the use of current (and flawed) ERTs, they do not provide principled reasons to constrain future systems that might work as intended. This exposes a conceptual and normative gap that the rest of the dissertation is designed to address.

The second part of the thesis, developed in **Chapters 3** and **4**, turns to emotional life itself to clarify what ERTs aim to track and what might be at stake in making emotional life fully legible.

Chapter 3 focuses on emotions and adopts an integrative perspective that identifies their central functions in social and moral life. It revisits historical debates about the place of emotion in morality through the lens of recent reinterpretations, emphasising emotions' normative significance for moral perception, deliberation, and motivation. It then surveys biological and evolutionary, cognitivist, and social constructionist perspectives, drawing out their shared functional commitments. On this basis, the chapter characterises emotions as evaluative, dynamic, and socially embedded phenomena that orient attention, shape practical and moral reasoning, and support interpersonal understanding. Tracking one's emotional states thus also implies tracking one's values, priorities, and unfolding moral commitments.

Chapter 4 shifts the focus from emotions as intrapersonal states to emotional expressions as the primary medium through which affective life becomes socially legible. It positions them as integral components of the informational infrastructure of social life, following two claims: (1) nonverbal affective communication constitutes a central channel of everyday social interaction, and (2) emotional expressions play this role because they function as defeasible yet generally reliable sources of information on which emoters and perceivers rely to coordinate action, manage expectations, and regulate social relations. The chapter then turns to emotional expressions in technosocial contexts. It argues that emotional expressions exhibit a bifunctionality under conditions of technological mediation. For human perceivers, emotional expressions function as communicative instruments through which social coordination, negotiation, and norm regulation are achieved. For ERTs, by contrast, the same expressions are treated as diagnostic data points used to infer, classify, and predict inner affective states. The chapter sketches a revised design-based account which conceives emotional expressions as relatively stable patterns of behavioural and physiological change that are designed, biologically, culturally, or conventionally, to render affective and motivational states recognisable rather than merely perceptible. This framework preserves a descriptive and function-first account of

emotional expressions while acknowledging that human and technological ‘perceivers’ take distinct epistemic routes to construct their own interpretation. From these distinct functional orientations emerges the idea that technology-based recognition may collapse the socially constructed opacity that currently sustains social interactions, limiting human inferences.

The third part of the dissertation develops the core (normative) contribution: the concept of emotional privacy and its place within informational privacy theory. **Chapter 5** begins by reviewing the evolution of privacy theory from early individualistic accounts through relational and social accounts to contemporary informational approaches. It shows how privacy concepts have repeatedly been adapted in response to new technologies and uses the recent emergence of mental privacy in the context of neurotechnology as a precedent for the kind of tech-induced gap that ERTs seem to create in our (normative) conception of privacy. To fill this gap, the chapter builds on the first two parts of the dissertation and on recent work in privacy theory to introduce a novel account of emotional privacy, treated as a subconcept of informational privacy. This conception of emotional privacy is presented as a dual-level structure. Actual Emotional Privacy concerns particular situations in which an individual can manage how legible their emotional states are to different audiences. Standing Emotional Privacy, in contrast, concerns the ability to maintain expressive discretion across time and contexts. The thesis argues that it is this standing dimension that is crucial for understanding what is at stake with fully functional ERTs. The analysis then demonstrates that emotional privacy protects expressive discretion and that such discretion is a constitutive condition for self-presentation, for rule-bound norms of expressiveness, and for the regulative and cohesive functions of emotional expression.

The final chapter of the thesis asks whether emotional privacy should be protected by a moral right and what difference this would make for the ethical evaluation of ERTs. **Chapter 6** defends the view that emotional privacy warrants recognition as a distinct moral right, demonstrating that it cannot be simply reduced to Control-based or Access-based moral rights to informational privacy or fully captured by Contextual Integrity. The argument demonstrates that technology-based emotion recognition is not a simple extension of ordinary mindreading but a different kind of recognition. When faced with fully functional ERTs, individuals would not only lose Actual Emotional Privacy, but they would risk losing the ability to exercise expressive discretion altogether. The chapter distinguishes acceptable losses of emotional privacy from wrongful ones through the notion of defeat, which marks the point at which

Standing Emotional Privacy is lost. To make these stakes concrete, the thesis develops three technomoral scenarios set at the individual, interpersonal, and societal levels. Across these scenarios, fully functional ERTs undermine practices of self-presentation, reduce the opacity that is necessary for the negotiated management of social roles, and disrupt the social functions of expressive norms in public life. The harms identified arise not only from illegitimate data flows or the exposure of affective content, but from a structural reorganisation of the expressive and epistemic fabric of social life. The chapter concludes by embedding the proposed moral right to emotional privacy into established ethical assessment frameworks. It shows that recognising emotional privacy alters the weight assigned to privacy in certain trade-offs within regulatory decision-making and provides firmer grounds for limiting or conditioning the deployment of affective technologies.

Through the chapters, the thesis offers a framework for understanding and evaluating the disruptive transformations that fully functional affective technologies could bring. 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. Its central conclusion is that 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.