



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).

1. General introduction

This opening chapter prepares the conceptual and methodological ground for the dissertation. After a preamble that frames the central philosophical problem (§1.1), I discuss the general assumptions underlying this work, the approaches chosen, the methodological choices (§1.2), and I conclude with an overview of the chapters of the dissertation (§1.3).

1.1. Preamble

Suppose governments, private companies, strangers or close relatives could have access to your inner emotions via a particular technology. Suppose law enforcement could use it for surveillance, policymaking, and to tailor public information campaigns; private companies could use it for hiring, insuring, or marketing purposes; strangers and close relatives could use it as a way to understand you better. Dystopia? Science fiction? Not exactly.

In our world, where technosocial contexts seem to be our new *zeitgeist*, peering into people's inner emotional life is something that an emergent technology has been experimenting with. Emotion recognition technology (ERT) uses AI algorithms to identify, collect, and process data extracted from the human face, body, movements, voice, gait, brain activity, and more, in the hope of making relevant inferences about emotional reactions, actions, motivations, and intentions. ERTs and related affective-computing systems have already been implemented across diverse contexts, including targeted advertising that analyses faces for negative and positive expressions (Affectiva; iMotions); emotional contagion on diverse social platforms (Kramer, Guillory, & Hancock 2014); wearables that track moods (Empatica); cars that detect emotions in drivers (Smart Eye; Affectiva Automotive AI); and glasses that detect people's emotions throughout a conversation (Emteq Labs). There is keen interest in using emotion-tracking software in policymaking, commerce, healthcare, and robotics. The field is moving swiftly towards real-time, multimodal, context-sensitive systems capable of making increasingly fine-grained inferences about affective and motivational states.

However, despite their growing deployment, ERTs remain highly controversial, with critics raising concerns about their lack of scientific legitimacy, ethical implications, and potential to reinforce discriminatory practices (Barrett et al. 2019; Tcherkassof & Dupré 2021; Crawford et al. 2019).

Yet while most attention is directed at the problems created by their current limitations, a crucial problem remains underexplored, namely how our social and normative landscapes would change if ERTs actually worked as intended. To my knowledge, no anticipatory ethical assessments have focused specifically on the case of fully functional ERTs. This dissertation aims to fill this critical gap. Rather than assessing current ERTs as they stand, this dissertation examines the social implications that would arise if ERTs were to be fully functional. Thus, this thesis investigates a central and timely question:

If ERTs were to function as intended, would this affect social practices, relations, and normative expectations, and if so, to what extent?

1.2. Assumptions, approaches, and methodological choices

This thesis contributes to the broader endeavour in the philosophy of technology that examines how technological and digital innovations shape human experience, social structures, and our understanding of the world. Its central aim is to anticipate the societal and ethical implications of a new technology, namely fully functional ERTs, in order to clarify what is at stake and what kinds of risks and normative challenges their use may introduce. The thesis is anchored in philosophy of technology given its central question, but the resources it mobilises bring together philosophy of technology, philosophy of emotion, and privacy theory.

The structure of the thesis is articulated in three distinct parts, corresponding to each of the broad topics of (i) technology (where I focus on ERTs), (ii) emotional life (where I focus on emotions and emotional expressions), and (iii) the ethical analysis (where I focus on emotional privacy). I dedicate one chapter to the topic of technology, two chapters to the topic of emotional life, and two chapters to the ethical analysis, centred on privacy. I explain the methodology undertaken in each of these parts in §1.2.2, but before I do that, I will say a few words about the way this structure came about within my personal reflection.

1.2.1. The reflexive process behind the dissertation's structure

One of the great things about philosophy is that it allows one to escape the methodological iron grip in which other disciplines are often trapped. Yet the danger associated with that freedom is to veer too far and lose touch with reality. While some, like Hermann Hesse, would probably argue that there is nothing wrong with such detachment, I have always been motivated by the

idea that philosophical work can make a genuine impact in society. Thus, one of the central aims of this thesis is to exploit the methodological freedom of philosophy to move beyond debates that focus narrowly on the present technical capacities of ERTs, while remaining firmly grounded in real-world concerns.

I began the thesis by mapping the technical landscape of ERTs. I had already examined existing (and thus inevitably flawed) systems in my master's thesis (Prégent 2021), but for this dissertation I conducted a second, extensive literature review. This revealed a striking series of recent breakthroughs in affective computing, which in turn seem to fuel the continued eagerness of both private and governmental actors to develop and deploy ERTs. While trying to understand the most likely path that research would take towards this development, I realised that the field of affective computing, despite its apparently strict methodologies, was drawing on anything and everything it could find in the study of emotions and emotional expressions to ground their own research. This meant often taking work out of context, overgeneralising claims made, and even relying on work that had been heavily criticised or disproven. This led me to question the true object of ERTs, which we often refer to as “emotions” or “inner affective states”. This is the moment I turned to the second part of this thesis.

I wanted to arrive at an understanding of emotions and emotional expressions that would be both sufficiently general and adequately tuned to the ambitions of ERTs. During my investigation, I realised that many theories which appeared, at least a priori, to be in radical opposition to one another (mostly due to their claims regarding the nature of emotions and emotional expressions) were in fact agreeing on many function-related aspects of emotions and emotional expressions. This motivated the integrative commitment I ultimately adopted, which I approached from a function-first angle. After spending more hours than I want to admit on this topic, I began reflecting on the fact that most research on ERTs focuses directly on the phenomenon that ERTs aspire to reach (viz. emotional states), but very few pause midway to consider the phenomenon that they *actually* reach (viz. emotional expressions). This made me realise that although emotional expressions are thought to be the expressive *component* of a given emotion, in social life, emotional expressions are much more than that and their causes do not limit themselves to emotional states but also include physical states and, more importantly for this research, social motives. As many researchers on ERTs were integrating under “emotion recognition” data related to stress, anxiety, but also dissatisfaction, and even things like “politeness”, I pursued the integrative approach and looked at transdisciplinary work on

emotional expressions. My aim was to find a way to define emotional expressions in a way that would allow both humans and technologies to draw their own inferences about an emoter's emotional states. To that end, I hypothesised and pursued the idea of a bifunctionality of emotional expressions. This bifunctionality offers an explanation on how 'meaningful' information can be captured through two different epistemic routes. From there, I began reflecting on the potential disruption that fully functional ERTs would pose to the functions that emotional expressions serve in social life. This was the turning point where I started reflecting on the third part of the dissertation.

Given these different epistemic routes, I wondered whether the information that was accessed through the technological route could put pressure on, or even severely disrupt, the current social structure of life, which relies heavily on the epistemic boundaries inherent to the "human route" of affective recognition and interpretation. This would imply that ERTs could provide information that were previously not available to ordinary perceivers under similar circumstances. Because emotional expressions are a type of information that is inherently social, I turned to both informational and social privacy, which significantly overlap in contemporary work. Using the methodological tools of conceptual engineering, I re-engineered the concept of emotional privacy, treated here as a subtype of informational privacy. It is a subtype of informational privacy because its object remains a kind of information, namely affective information. Emotional privacy is thus introduced as a conceptual addition to the privacy cluster, one that allows us to grasp more fully the normative (privacy-related) stakes of technologies such as ERTs. I argue that ERTs have a particularly disruptive potential with respect to the functions that emotional expressions serve in social life. This disruption deserves special attention not only because ERTs target emotional states that we would often prefer to keep private, but also because detecting such states risks undermining the core functions of voluntarily produced, and often conventionally crafted, emotional expressions on which social life relies.

As a result, another central aim of this thesis is to encourage future discussions on affective AI and privacy to adopt a broader conception of "emotional life" and to recognise the disruptive potential of ERTs for the structure of social life.

I now briefly sketch the main methodological choices underlying each of the three parts.

1.2.2. Methodological choices across the three general parts

Part one: The technology

Ethical reflection on emerging technologies faces a familiar methodological challenge known as the Collingridge dilemma (Collingridge 1980). It posits that, with regard to new and emergent technologies, we must choose between two evaluative stances: we must either (i) evaluate a technology in its early stage, or (ii) evaluate it later in its development. If we choose the former, we presumably still have reasonable control over its design but likely lack sufficient knowledge of its social and moral implications. If we choose the latter, and evaluate it later in its development, these social and moral implications have likely become clearer through widespread adoption, yet at this stage the technology is often so embedded and entangled in the social, economic, or political fabric that redirecting or constraining it becomes extremely difficult.

Sociotechnical experimentation and responsible innovation, for instance, focus on this latter stage. They aim to monitor technologies and adjust governance practices once systems are already in place (van de Poel 2013; Owen, Bessant, & Heintz 2013). However, such approaches inevitably remain backward-looking, responding to harms or unintended consequences only after the technology has already reshaped social practices. As a result, they are ill-suited to prevent harms that a more forward-looking evaluative approach, such as anticipatory ethics, might have identified long before actual deployment.

Anticipatory ethics focuses on the former stage and seeks to examine the possible future impacts of emerging technologies while intervention is still possible, bringing ethical analysis to the forefront of technological design and development (Swierstra & Rip 2007; Brey 2012; Vallor 2016).¹ However, working this early in a technology's life cycle means that anticipatory approaches risk drifting into purely speculative scenarios that bear little relation to real technological trajectories (Kudina and Verbeek 2019; Nordmann 2007). The main concern is that these analyses risk developing unrealistic scenarios that lead them to focus on theoretical questions that have little to no practical significance.² A responsible anticipatory methodology

¹ It might be worth noting that a third approach to the Collingridge dilemma has recently been proposed by Kudina and Verbeek (2019). Kudina and Verbeek propose the “ethics from within” approach, a type of technological mediation that focuses on the problem of value dynamism in new and emergent technologies.

² In other words, to borrow a popular expression from Québec, they risk ending up “pelleter des nuages” (shovelling clouds).

thus requires that its analysis be empirically grounded whenever possible, so as to understand and not depart from a technology's trajectory.

I believe that neither of these approaches is sufficient in itself and that both types of approaches are important for responsible innovation. Having said that, this thesis adopts and contributes to the former stage of the development of a new technology, following anticipatory ethics. Consequently, in the remainder of this paragraph I now briefly explain how I have sought to mitigate the risks that come with anticipatory approaches, particularly the risk of drifting away from real-world issues.

While the ultimate aim of this thesis is to contribute to mapping the ethical risks associated with the possibility that ERTs work as intended, I pre-empt some of the main methodological pitfalls of anticipatory ethics by drawing the prospective scenario of fully functional ERTs on how present developments appear to be moving. This entails a detailed reconstruction of their *current* sociotechnical trajectory across technical, social, economic, and regulatory contexts. In this way, I aim to ensure that the framing of the ethical problem is empirically anchored rather than completely detached from actual developments. This provides a plausible hypothetical scenario of fully functional ERTs grounded in existing research, commercial applications, and regulatory debates. Additionally, I was careful to not ground the ethical analysis too tightly on the empirical predictions regarding how, exactly, such technology would work. Even though I do make some assumptions throughout the thesis, the ethical analysis and its normative conclusions are only grounded on the high-level assumption that these AI systems work.

Part two: The emotional life

Part two turns to the object of ERTs, namely the emotional life. The methodological stance in these chapters is transdisciplinary and function-first. I bring together work from the philosophy of emotion, moral and social psychology, neuroscience, anthropology, and affective science in order to develop a joint account of emotions and emotional expressions. Rather than treating these disciplines as competing sources of authority, I use them as complementary lenses that illuminate different dimensions of affective life.

Chapter 3 adopts what I call an integrative perspective on emotions. This is also a function-first methodological lens that explains emotions in terms of the roles they play in perception, practical reasoning, social coordination, and normative life. It integrates insights from biological

and evolutionary approaches, cognitivist and evaluative approaches, and social constructionist approaches to make sense of how emotions structure attention, confer salience, guide action, and sustain interpersonal relations. The chapter also employs historical and conceptual analysis of major traditions in moral philosophy (e.g. moral rationalism, virtue ethics, sentimentalism, and contemporary hybrid accounts) in order to situate emotions within debates about moral judgement and motivation. The aim is not to resolve foundational debates about the nature of emotion, but to extract a multi-level functional picture that can be used in Chapter 4 and the later ethical assessment of ERTs.

Chapter 4 continues this function-first approach by focusing on emotional expressions and affective communication. Methodologically, it develops a conceptual overview of emotional expressions, analysing their properties, causes, and varieties, and it proposes to see emotional expressions along a spectrum, which includes natural, semi-natural, semi-conventional, and conventional expressions. It then endorses a probabilistic and pragmatic view of the informational role of emotional expressions. On this view, expressions are treated as a reliable though defeasible source of information in virtue of probabilistic regularities that are shaped by context, norms, and background knowledge. Building on this, the chapter advances a modified design-type account of emotional expressions. This allows me to distinguish the functional roles emotional expressions play within human and ERT-based recognition while preserving a unified account of emotional expressions.

Across both chapters, the analysis is structured by a distinction between intrapersonal and interpersonal levels. Chapter 3 focuses on the intrapersonal dimension of emotions as states that shape deliberation and action readiness within the individual. Chapter 4 focuses on the interpersonal dimension of emotional expressions as elements of affective communication that coordinate behaviour, regulate social expectations, and support mutual understanding. Taken together, these methodological choices yield an integrative and functionalist account of emotional life that prepares the conceptual ground for the subsequent ethical analysis of ERTs from a privacy perspective.

Part three: The (ethical) analysis

Once the technological and affective foundations are in place, the third part of the thesis turns to the ethical analysis proper. ERTs raise many ethical concerns, from creating new possibilities for manipulation and discrimination to posing threats to privacy and democratic practices. In

this third part, however, I intentionally restrict the ethical analysis to the informational privacy dimension. I do so because it is within the informational privacy paradigm that fully functional ERTs appear to introduce a new and structurally distinct kind of harm, namely the defeat of expressive discretion.

The guiding question for this part is not simply whether ERTs are permissible, but whether existing informational privacy concepts and corresponding moral rights are suitably equipped to register and protect against the kinds of harms that arise when affective life becomes fully legible to technological systems. The ethical analysis deliberately brackets issues such as bias and unreliability so as to be able to isolate the privacy-related harms that stem from the very possibility of continuous affective legibility. Methodologically, I divide this task into a descriptive-conceptual stage (Chapter 5) and a normative stage (Chapter 6). In Chapter 5, I adopt a problem-driven, cluster-based form of conceptual engineering. Rather than treating privacy as a fixed concept, I treat it as a conceptual cluster that has historically been adapted in response to new technologies. Using this cluster-based approach, I identify a tech-induced conceptual gap. I thus proceed by asking what conceptual resources would be needed to fill this gap. I take inspiration from the recent introduction of mental privacy in response to neurotechnologies and treat it as a useful case study for this kind of conceptual adaptation. I then draw on it in developing emotional privacy. Emotional privacy is introduced as an additional subconcept to the informational privacy cluster. Once this concept is in place, I turn to the social and moral stakes raised by technologies that target emotional expressions and affective communication and the normative task of determining whether, and in what sense, emotional privacy should warrant normative protection.

To answer that latter question, the normative stage relies on technomoral scenarios, structured across individual, interpersonal, and societal levels of interaction, to examine how the defeat of expressive discretion manifests in different social roles and institutional settings. These scenarios then feed into Contextual Integrity and ALTAI-style ethical assessments, which are used not to derive a complete ethical verdict on ERTs, but to test whether recognising a moral right to emotional privacy can genuinely change the overall verdict of an ethical assessment on fully functional ERTs.

1.2.3. Transdisciplinary terminology and conceptual trade-offs

While being firmly anchored within the philosophical field, this thesis also relies on transdisciplinary work from psychology, anthropology, affective sciences, neuroscience, and computer science. This has enabled me to develop a rich, detailed, and nuanced understanding of the concepts and phenomena at stake from diverse points of view, but it has also brought specific challenges.

The main drawback of relying on such diverse literatures is that it has forced some terminological “trade-offs”. Philosophically meaning-laden words, such as “truth”, “knowledge”, “emotions”, and “happiness”, have, at times, had to be relaxed or partially bracketed in favour of some of their counterparts in other disciplines. Without this adjustment, it would have been impossible to engage seriously with these fields’ own operative understandings of the concepts they invoke. I therefore engage with these disciplines’ understandings of our shared concepts, subjecting their interpretations and the theoretical commitments they embody to philosophical analysis and, where needed, criticise them (for example, the tendency in affective computing and computer science to over-reduce, or to include too wide a range of phenomena among those that count as an “emotion”).

1.3. Overview of the chapters

Below I offer an overview of the five chapters that make up the body of this dissertation.

Chapter 2 provides a comprehensive overview of ERTs in their social, technological, and political contexts, establishing the groundwork for the thesis’s normative inquiry centred around the possibility of “fully functional ERTs”. It first traces the historical development of affective computing, from its roots in cognitivism to the emergence of multimodal ERTs, showing how technological progress has moved beyond static, face-based systems towards increasingly context-sensitive and adaptive models. It then surveys the current ERT market covering embodied agents, self-tracking apps, passive surveillance tools, border-control lie detectors, and aggression detection systems, highlighting their multiple functions across commercial, healthcare, educational, and security domains. It also examines technical limitations and pinpoints key challenges, such as noise, indeterminacy, and the absence of ground truth, while acknowledging ongoing advancements that could soon yield “fully functional” ERTs capable of informing high-stakes decisions. The chapter also reviews critical scholarship and political pushback, focusing on biases, the contested scientific foundations of ERTs, and concerns about

their discriminatory potential. Finally, it analyses current international regulation, focusing on the EU AI Act's treatment of ERTs, identifying two key shortcomings: its narrow focus on technical unreliability and its failure to provide a normative justification for restricting even highly accurate future systems. The chapter concludes by positioning the thesis as an anticipatory ethical analysis aimed at addressing this gap.

Chapter 3 turns to the phenomenon of emotion itself, examining its nature, its role in (moral) decision-making, judgement, and action, and its functioning at the intrapersonal level. It begins by distinguishing emotions from neighbouring intentional states such as beliefs, desires, and judgements, and emphasises their multifaceted nature as evaluative, motivational, and phenomenological phenomena embedded in social life. The chapter then examines the historical debate on emotions in morality: while moral rationalists (Plato, Kant) sought to exclude or control emotion, their indispensability for moral reasoning emerged in contrast; virtue ethics (Aristotle) and sentimentalism (Hume) instead foregrounded their constitutive role in moral perception and motivation. Contemporary theories (Nussbaum, Tappolet, Greenspan, Ben-Ze'ev, and Blum) further affirm emotions as epistemically significant and morally salient.

Building on this, the chapter surveys three major perspectives, namely, biological/evolutionary, cognitivist, and social constructionist, treating these perspectives as complementary lenses rather than rival theories. From their points of convergence, it extracts a picture of emotions as embodied, evaluative, culturally shaped, and socially embedded processes that are adaptive, normative, and dynamic. Yet the chapter also stresses that to speak of “the functions of emotions” without further specification can be misleading as different emotions fulfil different roles, and even a single emotion type can serve distinct functions depending on context, background knowledge, and internal tensions. On this basis, the chapter adopts an integrative perspective on emotions that understands emotions through the roles they play in structuring perception, orienting attention, guiding action, and mediating interpersonal relations. This approach underpins the thesis's later normative evaluation of ERTs, highlighting how tracking affective life also partly tracks one's decisions, values, and action tendencies. If emotions can be measured and understood with sufficient precision, their instrumentalization becomes a practical possibility. Yet a full account of emotional life also requires examining its outward dimension, which is the focus of the next chapter.

Chapter 4 shifts the analysis from the intrapersonal to the interpersonal domain by focusing on emotional expressions (EEs) as the primary medium through which affective states become socially legible. It argues that for ERTs to pose a substantive normative challenge, two conditions must hold: first, nonverbal affective communication must play a central role in everyday interaction; second, EEs must function as reliable sources of information on which emoters and perceivers depend to sustain other vital social functions. This chapter shows that these two conditions are met as EEs are primary elements of the affective communication channels on which interactions depend, sustaining coordination, managing expectations, and regulating social life. The chapter begins by providing a theoretical grounding of EEs, identifying their core properties, causal sources (including bodily states, emotions, and social motives), and mapping a spectrum from natural through semi-natural and semi-conventional to fully conventional EEs.

It then situates EEs within affective communication channels, adopting a probabilistic-pragmatic approach according to which emotional expressions convey meaningful and reliable affective information even though perceivers need to partly rely on social norms and background knowledge to make sense of it. Building on this, the chapter develops a functionalist account of EEs that positions itself between both reductive naturalism and pure perceiver-dependent accounts. In technosocial contexts, this framework underwrites a bifunctionality of EEs: for humans, EEs operate as communicative acts that structure and regulate social life; for ERTs, they become diagnostic data points for inferring inner states. In doing so, the chapter clarifies how EEs can, through two different epistemic routes, simultaneously sustain human social practices and feed into automated recognition systems, setting the stage for the later privacy-focused ethical analysis.

Chapter 5 develops the conceptual and theoretical foundations for the concept of emotional privacy, positioning it as a constitutive condition of social life rather than a mere shield from it. Building on the functionalist account of emotions and emotional expressions developed in Chapters 3 and 4, it reconceptualises privacy in affective life as the condition under which emotional expressions can fulfil their social roles. It first traces the historical evolution of privacy from individualistic to relational and social accounts and surveys contemporary informational privacy scholarship in order to show how privacy concepts have repeatedly been adapted in response to technological change. Using the recent emergence of ‘mental privacy’ in response to neurotechnologies as a precedent, the chapter argues that fully functional ERTs generate a

similar tech-induced conceptual gap in the affective domain that the informational privacy cluster currently cannot fill.

On this basis, Chapter 5 introduces and defines emotional privacy as a subconcept within the informational privacy cluster. It distinguishes between *actual emotional privacy*, which obtains when an agent enjoys expressive control in particular situations, and *standing emotional privacy*, which obtains when an agent retains the expressive ability to manage affective legibility over time. This two-level structure captures the idea that when it comes to emotional life, privacy concerns not only particular episodes of access but the background ability to modulate what one's expressions count as in context. The chapter thus argues that emotional privacy aligns with the Control-based perspective of informational privacy, even though it departs from standard Control-based accounts by showing that the obtention of standing emotional privacy, and not actual emotional privacy, is what matters for expressive discretion.

The remainder of the chapter examines how emotional privacy functions within affective communication channels. It first analyses expressive conventions and affective leakages, and then turns to privacy mechanisms, distinguishing between mechanisms of selective disclosure and mechanisms of concealment. These mechanisms operate by modulating the degree of affective opacity in given interactions, across audiences and contexts. The functional value of these mechanisms is then traced across three levels of social life: at the individual level, emotional privacy underwrites self-presentation; at the interpersonal level, it enables the management of role plurality and adherence to role-bound norms of expressiveness; and at the societal level, it supports social regulation and cohesion. The chapter concludes that (standing) emotional privacy is an integral part of the infrastructure of social life, and that its loss would disrupt not only individual interests but also the social functioning of affective communication. This grounds the normative inquiry undertaken in Chapter 6, which asks whether emotional privacy should warrant normative protection, and if yes, how recognising such a moral right would reshape the ethical assessment of ERTs.

Chapter 6 makes the normative turn of the thesis by developing and defending a moral right to emotional privacy and exploring its implications for the ethical assessment of fully functional ERTs. Building on the descriptive account in Chapter 5, it first articulates a normative account of emotional privacy and formulates conditions under which this right is realised or violated. On this view, emotional privacy is violated when technologies defeat individuals' expressive

discretion. The chapter then argues that this right cannot be reduced to, or fully captured by, existing normative frameworks, including Control-based and Access-based moral rights to informational privacy, and Contextual Integrity. This distinctiveness is further clarified through an analysis of the differences between human and technology-based emotion recognition, showing that fully functional ERTs restructure the affective and expressive fabric of social life rather than merely extending ordinary mindreading. To make the argument concrete, the chapter employs three technomoral scenarios, staged at the individual, interpersonal, and societal levels, which illustrate how the continuous legibility of affective life undermines practices of (i) self-presentation, (ii) role plurality and adherence to role-bound norms of expressiveness, and (iii) social regulation and cohesion. Finally, the chapter tests the proposed moral right by embedding it in Contextual Integrity and ALTAI-style ethical assessments, showing how recognising emotional privacy may change the moral evaluation in favour of privacy, particularly in trade-offs involving ERTs and other similar technologies. In doing so, Chapter 6 demonstrates that emotional privacy is a helpful addition to the informational privacy cluster when evaluating current and future affective technologies that pose a distinctive threat to individuals' expressive discretion.

Together, these chapters render intelligible the central claim of the thesis: that biases and discrimination caused by flawed ERTs only represent the tip of the iceberg when it comes to understanding and mapping the social and ethical impact of ERTs. Tech-induced change is not inherently problematic, but forecasting potential risks and challenges is essential to guide the development and use of these new technologies, as part of a collective effort to remain aware, reflective, and normatively informed regarding the changes that lie ahead. By combining an empirically informed anticipatory analysis of fully functional ERTs (part one), a function-first integrative account of emotions and emotional expressions (part two), and developing the concept of emotional privacy (part three), the dissertation contributes to contemporary debates in the philosophy of technology, the philosophy of emotion, and privacy theory, and offers a normative framework for future philosophical, ethical, and regulatory work on affective AI systems.