Emotion Inferences in the Workplace and Healthcare: Workers' and Patients' Emotional Privacy Judgments and the Relative Influence of Contextual, Socio-demographic, and Individual Privacy Belief Factors

ANONYMOUS AUTHOR(S)

The increasing use of emotion AI in workplaces and healthcare has raised ethical and privacy concerns. We conducted a factorial vignette survey with a U.S. nationally representative sample (n=300) and a sample representing minoritized groups (n=385) to investigate workers' and patients' privacy judgments concerning their emotions inferred by these technologies. Participants judged 56 scenarios involving emotion inferences across workplace and healthcare settings that varied by data input (image/video, speech/text) and purpose (e.g., diagnostics, harm prevention), and reported their socio-demographic information and privacy beliefs. Findings surface the relative importance of contextual, socio-demographic, and individual privacy belief factors on workers' and patients' generally unfavorable emotional privacy judgments, with notable sample-level differences. Our research underscores the significance of purpose and minoritized populations to privacy theory, research, and policy; and demonstrates the need for policies that recognize emotional information as a sensitive data category and advocate inference minimization principles.

CCS Concepts: • **Human-centered computing** → Empirical studies in HCI.

Additional Key Words and Phrases: AI, emotion AI, emotion recognition, affective computing, workplace, mental health

ACM Reference Format:
Anonymous Author(s). 2018. Emotion Inferences in the Workplace and Healthcare: Workers' and Patients' Emotional Privacy Judgments and the Relative Influence of Contextual, Socio-demographic, and Individual Privacy Belief Factors. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 55 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Emotion AI technologies process multimodal (and often identifiable) data sources (e.g., text, speech patterns, facial expressions) to infer human affect and sometimes respond with simulated emotional sensitivity [48, 116, 151, 158]. Throughout this paper, we use "emotion AI" to refer broadly to technologies that process identifiable input data to infer emotional or affective states, regardless of technical taxonomy. Once a niche area of research, these technologies are now increasingly embedded in workplace and healthcare settings [27, 32, 73, 90, 116, 119]. Deployments promise improved safety, performance, and well-being [27, 73, 90, 116, 119]. Yet they also facilitate unprecedented flows of emotional and affective data—flows that can reinforce power asymmetries, reproduce demographic biases, and enable surveillance and other misuses of sensitive emotional information, with effects that erode institutional trust and harm those subject to the technology [130, 145, 189, 192]. For workers and patients whose jobs or care may hinge on opaque inferences generated by these technologies, the stakes can be profound.

Understanding when and how privacy is transgressed is essential to grasping the social impacts of emerging technologies and mitigating their harms [37, 159]. Yet empirical knowledge on how workers and patients—the individuals most exposed—judge the appropriateness of emotion AI data flows, and how those judgments vary with context, social position, and privacy beliefs, remains scarce. In its absence, technology research, policy, and practice risk privileging dominant norms while overlooking the heightened vulnerabilities of minoritized groups. **RQ:** *What is the relative influence of contextual, socio-demographic, and individual privacy belief factors on workers' and patients' emotional privacy judgments of emotion AI data flows?*

To answer this question, we designed a factorial vignette survey based on Helen Nissenbaum's theory of contextual integrity, which normatively justifies data flows when they uphold the legitimate goals of the context and serve its broader social ends [124]. To conceptualize emotional privacy as the appropriateness of emotional information flows, we structured vignettes by fixing contextual integrity's five canonical parameters: information type, subject, sender, recipient, and transmission principles. Guided by the principle of purpose limitation—which restricts data use to specific, legitimate aims [59, 74]—we systematically varied vignettes by 14 emotion data *purposes* (e.g. safety, diagnostics, performance management) and two *input* modalities (image/video vs. speech/text). Participants rated their comfort across 56 scenarios in total (2 contexts x 2 inputs x 14 purposes). We also collected *socio-demographic* factors and individual *privacy beliefs* (e.g., institutional trust, perceived sensitivity of emotional information) to model their influence on privacy judgments. Recognizing that nationally representative samples may obscure the privacy needs and vulnerabilities of underrepresented groups [114], we conducted the study across two U.S. adult samples: a nationally representative cohort by race, sex, and age (n=300) and a targeted oversample of minoritized participants by race/ethnicity, gender, and mental health status (n=385). We analyzed cohorts separately to reveal patterns that pooled, weighted analyses might miss. [1] Our results yield three key insights for emotional privacy theory, system design, and governance:

1. Purpose is a dominant, context-specific driver of privacy judgments. Holding contextual integrity's actor, attribute, and transmission principle parameters constant, varying the stated *purpose* of an emotion AI flow shifts mean comfort by −7.9 to +7.0 points—the largest swings observed. Purpose shows an interdependent effect: its direction and magnitude vary with the institutional goals of each domain, with some purposes producing the strongest effects across the model. In contrast, *input* modality shows a consistent main effect across contexts and cohorts: replacing image/video with speech/ text raises comfort by +2–5 points, reflecting generalized discomfort with facial emotion analytics.

Employment. Flows supporting the workplace's social mission—keeping workers safe, cared for, and productive—raise comfort: risk-of-harm predictions (≈ +7), group mental-health monitoring (+2.6/+2.2), and automated acute support (+1.7/+1.9). Conversely, evaluative flows importing clinical diagnoses or expanding individual surveillance—early medical diagnoses, individual mental health inferences, and performance scoring—lower comfort (−1.3 to −3.7). While these flows might, in principle, aid productivity and care, participants appear to weigh disclosure risks to employers more heavily, recognizing the power asymmetries such flows may intensify—thus contravening employment's higher-order social goals (e.g., dignity, fair treatment [9]).

Healthcare. Early neurological screening is the only purpose rated positively in both samples (+2.2/+3.5), aligning with healthcare's aim of improving clinical outcomes. Yet other clinical purposes—early mental health diagnosis, individual mental health inference, and automated interventions—register the strongest negative effects (−3.5 to −7.9), suggesting that granting machines interpretive authority over emotional states heightens vulnerabilities and undermines bodily and decisional autonomy—core to healthcare's contextual integrity [177].

[1] Open-ended responses on perceived benefits and risks are published elsewhere [41, 143]. We reference those findings only where they contextualize our quantitative results.

These purpose-specific variations reflect contextual integrity's normative claim: a flow is judged appropriate when its purpose furthers the context's institutional goals and the broader social ends they serve, and inappropriate when it strains or distorts those ends. Our findings therefore support governance efforts to bind the flow of emotion inferences to narrowly articulated, context-serving purposes and apply purpose limitation principles by default.

2. Socio-demographic variation shapes emotional privacy judgments in context. Our dual-sampling approach highlights differing privacy judgments between representative and minoritized cohorts. Across both employment and healthcare settings, participants in the minoritized cohort tended to follow the same directional trends as the nationally representative cohort—but with magnified effect sizes, both positive and negative. These differences suggest heightened perceived susceptibility to emotional inferences and greater judgment intensity, consistent with the idea that position-related *vulnerability* shapes privacy expectations [114].

Importantly, divergences emerged at the purpose level. For example, in employment, early diagnosis for mental illness had a more negative effect in the minoritized sample (−2.5 vs. −1.3), as did neurological disorder screening (−1.7 vs. +0.5). These divergences were also context-specific. For instance, in healthcare, early mental health diagnosis was rated less negatively by the minoritized cohort (−0.5 vs. −2), while neurological screening was rated more positively (+3.5 vs. +2.2). These results suggest that participants more acutely attuned to systemic disparities (e.g., in care access, stigma) may evaluate flows through a different lens of contextual appropriateness—recognizing, for example, how a given use might support or undermine a context's broader social ends. This position-sensitivity is further evident in the one statistically significant reversal of effect direction: identifying patients in need of support in healthcare, which was rated negatively by the representative cohort (−1.2) but positively in the minoritized cohort (+1.3). Such divergences highlight how differing lived experiences inform privacy judgments about whether a data flow upholds or violates contextual integrity. These divergences underscore how lived experience shapes judgments of contextual appropriateness.

Concerning socio-demographic factors, while not all effects were statistically significant, patterns by race, gender, mental health status, and education were nonetheless illuminating, wherein we observed patterns consistent with position-related vulnerability. In both contexts, Black participants and those without a Bachelor's degree tended to report greater comfort, suggesting heightened sensitivity to flows that promote well-being and dignity. Meanwhile, gender minorities and participants with mental health histories exhibited sharper negative responses in healthcare, highlighting where emotion AI systems may exacerbate existing vulnerabilities. These findings underscore the need for demographic sensitivity in the design and governance of emotion AI, and caution against assuming nationally representative samples capture the full range of emotional privacy concerns.

3. Trust and perceived sensitivity are decisive belief factors. Institutional trust and perceived sensitivity of emotional data strongly influence comfort judgments. Each one-unit increase in trust, or decrease in perceived sensitivity, shifts comfort by 0.4 to 0.5 points, comparable to many mid-range purpose effects. Specifically, institutional trust increased comfort by +0.44 to +0.54 per scale unit, while perceived sensitivity decreased it by −.25 to −0.3. Notably, perceived sensitivity of emotional information varied by context and, in some cases, was even rated higher than traditionally recognized sensitive categories of data such as biometric, genetic, or union membership information. This underscores the need to treat emotional information as a first-order privacy concern. Because these are continuous variables, their cumulative effect may exceed that of any single contextual or demographic variable we tested. As trust varies widely across workplaces and healthcare settings, meaningful protections should therefore be built into systems by design—not deferred to assumed institutional goodwill—and governed according to the heightened sensitivity of emotional information.

By grounding emotional privacy as a normative judgment shaped by contextual goals, individual beliefs, and position-related vulnerabilities, this study extends contextual integrity by incorporating diverse participant perspectives and empirically testing the influence of purpose—alongside fixed contextual integrity parameters—on privacy judgments. In doing so, we offer both a theoretical refinement and an actionable model for evaluating the acceptability of emotion AI systems within the contextual integrity framework. These insights lay the foundation for designing and governing emotion AI technologies that respect autonomy, support dignity, and advance the broader social ends these systems claim to serve. At the same time, they surface a critical challenge: ensuring that the full socio-technical pipeline—from purpose specification to transmission constraints and system safeguards—consistently upholds these normative commitments across specific deployments and contexts. Our findings also offer empirical support for recent regulatory developments, such as the EU AI Act, and provide a model for anticipating future governance needs. In the sections that follow, we elaborate this framework, present our empirical findings, and discuss implications for the responsible development, deployment, and regulation of emotion AI grounded in heightened safeguards for emotional information, purpose-aware design and governance, and sensitivity to contextual and positional vulnerabilities in privacy research and practice.

2 BACKGROUND

Despite growing deployments of emotion AI in the workplace and healthcare, the privacy implications of these technologies remain poorly understood—particularly from the perspective of those most affected. While calls for empirical attention to privacy in technologies handling emotional data are growing, particularly in applications of AI to the workplace [104, 145, 191] and healthcare [23, 130, 153, 189], two persistent gaps limit progress. First, there is a structural gap: the actors designing, deploying, and evaluating these technologies often lack insight into the situated norms and vulnerabilities of the individuals over whom they hold power [47, 112]. Second, there is a conceptual gap: emotional privacy remains empirically under-theorized and thus difficult to operationalize [145], with limited empirical attention to how emotional information flows are judged across contexts, identity characteristics, and belief systems.

The present study addresses both gaps by investigating how three interdependent dimensions—(1) contextual factors, (2) socio-demographic characteristics, and (3) individual privacy beliefs—influence emotional privacy judgments. Our analytic framework models the relative influence of these factors on workers' and patients' judgments of emotion AI data flows *alongside* the structural parameters of contextual integrity. In doing so, we build on contextual integrity theory to clarify what emotional privacy means in practice, identify the factors that shape its protection or violation, and inform governance efforts to align emotion AI with both human values and contextual demands. This section reviews literature motivating our inclusion of these factors as explanatory variables alongside contextual integrity parameters, with attention to how identity-based vulnerabilities can influence emotional privacy judgments in work and healthcare domains.

2.1 Contextual Factors Shaping Emotional Privacy Judgments

According to the theory of contextual integrity, privacy norms are shaped by the interdependent parameters of a given social context, including actors, attributes, and transmission principles. An information flow is judged appropriate when it aligns with these norms and supports the context's institutional goals [124]. What is acceptable in healthcare may not generalize to the workplace; these domains differ not only in place but in politics, conventions, and cultural expectations [124, 181]. Attending to the specific contextual configurations of employment and healthcare is thus essential for evaluating whether, and to what extent, emotional privacy is preserved or violated. Such analysis can reveal gaps between normative ideals and lived experience, enabling more socially responsive design and policy aligned with the values of those most affected by emotion inference technologies. In addition to measuring emotional privacy through contextual integrity's core parameters, this study examines the relative influence of two further contextual factors: the modality of data *input*, and the stated *purpose* of its use. These are particularly salient in emotion AI, where inferences are drawn from multimodal signals and often used for opaque or poorly justified ends [27, 95, 144].

2.1.1 Data Input. In both workplace and healthcare contexts, emotion recognition frequently relies on text, speech, and facial data, often in combination with additional contextual or biometric information [13, 27, 45, 93, 167]. The specific input modality may influence emotional privacy judgments, as privacy perceptions are known to vary across data types. Facial and bodily data captured through cameras and facial recognition systems raise concerns about visibility, identification, and biometric surveillance [14, 163, 188, 191]. Speech data collected via continuous microphones, such as those in smart speakers, elicit fears of ambient surveillance and constant monitoring [94]. Text-based inputs, including monitored emails and messages, prompt concerns about intrusions into private communication and intent inference [182]. In emotion AI, such data are not shared directly but processed to infer emotional states and then automatically disseminated. Under the theory of contextual integrity, these input modalities form part of the broader "sender" of emotional information [125].

Empirical studies support the idea that input type shapes privacy perceptions. Lee et al.'s qualitative study on mobile affective computing found sensor-specific concerns, with users worried that certain data types could expose personal traits and lead to profiling and surveillance [95]. Similarly, Zhang et al. showed that inferences about mental health from mobile data triggered privacy concerns that varied by data source and contextual framing [189]. These findings suggest that the input source used to generate emotion inferences may directly shape how workers and patients evaluate emotional privacy.

2.1.2 Purpose. The purpose for which information is collected and used is known to shape privacy perceptions—particularly when the stated purpose offers a personal or collective benefit [44, 51, 86, 115, 121, 142, 147, 184]. Individuals are often more willing to share sensitive information, including emotional or health-related data, when they perceive it as contributing to their own well-being [21] or advancing a broader social good [77, 95, 172]. However, purpose-driven framing can be leveraged by powerful actors to normalize surveillance and downplay privacy risks. In workplace contexts, for instance, employers increasingly frame monitoring tools as enhancing productivity or well-being, thereby encouraging participation while limiting dissent [5]. Similarly, in healthcare, optimistic narratives about digital tools can obscure underlying privacy threats [20]. While positive framing may mask certain risks, people remain concerned about their emotional privacy even when purported benefits are emphasized. Zhang et al. found that privacy concerns persisted in mobile mental health apps, despite framing these tools as beneficial [189].

These findings highlight the contextual salience of *purpose* in shaping emotional privacy perceptions. Indeed, U.S. privacy regulations such as the Health Insurance Portability and Accountability Act (HIPAA) embed purpose specificity as a normative constraint on sensitive data use [3, 62]. While purpose is not explicitly included among contextual integrity's five canonical parameters, applications of the theory often treat purpose as an optional transmission principle constraining data use [124]. Given that emotion AI systems infer emotions from diverse data inputs and for a range of ends, we expect emotional privacy judgments to be shaped not only by the structural features defined by contextual integrity, but also by the *purpose* for which emotion inferences are used. By empirically assessing the effect of purpose across fixed contextual integrity parameters, our study extends contextual integrity by measuring how purpose meaningfully contributes to emotional privacy judgments in context-specific ways.

2.2 Socio-Demographic Variation in Emotional Privacy Judgments

While contextual factors define the descriptive and normative boundaries of privacy within a given setting, individuals' socio-demographic characteristics shape how those boundaries are perceived, enforced, and contested.

Empirical work shows that privacy perceptions vary across socio-demographic dimensions such as education [21], race/ethnicity [21], and gender [19, 96]. Although the relationship between privacy perceptions and socio-demographic status remains understudied [89], research in privacy and HCI suggests that identity-based characteristics influence both one's exposure to surveillance and sensitivity to its harms [19, 67, 69].

2.2.1 Education. Recent Pew findings indicate that public concern over AI applications varies by educational attainment. Individuals with postgraduate education expressed greater concern about facial recognition by police, while those with a high school diploma or less were more concerned about AI-enabled misinformation detection and autonomous vehicles [137]. Similarly, Bhatia and Breaux found that individuals with doctorate degrees reported lower concern about sharing personal information than those with lower educational attainment [21].

2.2.2 Race/Ethnicity. Research shows that Black and Latine populations are afforded less privacy in U.S. society, in part due to the long-standing normalization of racialized surveillance practices [30, 39]. This structural disparity may contribute to privacy resignation and a diminished perception of risk, despite disproportionately high levels of vulnerability to privacy intrusions [67]. People of color may face heightened risks from emotion AI technologies, as studies have found that emotion recognition algorithms using speech, facial analysis, and natural language processing are often less accurate for people of color—increasing the likelihood of misclassification and harm [75, 139, 187].

2.2.3 Gender. Gendered surveillance also shapes privacy experiences and concerns. Women, who face disproportionate exposure to gender-based harassment [87, 155] and workplace monitoring [163], consistently report heightened privacy concerns in these contexts [10, 16, 57, 71, 176]. While research on the privacy perceptions of transgender and non-binary individuals remains limited, existing scholarship suggests that gender minorities have distinct privacy needs—shaped by elevated vulnerability to technological harms from exclusion and exposure [150, 179], as well as a heightened reliance on safe, supportive, and affirming technology-mediated interactions [68, 70, 98].

2.2.4 Mental Health Status. The privacy perceptions of individuals with mental illness warrant special consideration, as this population may be more vulnerable to the impacts of automated emotion inference. While emotion AI applications may offer mental health benefits in some cases [90], individuals with mental illness also face heightened risks from stigmatization, disability discrimination, and inaccurate inferences [119]. Research centered on the privacy perspectives of individuals with mental illness highlights a constrained choice architecture: despite recognizing personal privacy risks, individuals report feeling compelled to trade privacy for access to potentially beneficial mental health technologies, including those that rely on emotion inferences [23, 42].

These risks are magnified in employment and healthcare contexts. Workers with mental illness frequently hesitate to disclose their conditions due to fears of discrimination, damaged professional reputations, and lack of confidentiality. Disclosure decisions often involve weighing the potential benefits of accommodations against risks of stigma and exclusion [29, 50, 169]. Similarly, patients with mental illness report distinct privacy concerns about health information sharing—even in clinical settings—stemming from prior experiences of mental health-related mistreatment [153]. These concerns extend to mobile mental health applications, where participants express particular discomfort with the sharing of sensitive data types such as social interaction information—potentially due to lived experiences with isolation and vulnerability [189]. The emergence of emotion AI in digitized workplace and healthcare settings may compound these challenges. Emotion AI systems have roots in problematic efforts to pathologize affective difference, such as the stigmatization of emotional expression in autistic individuals [122]. These technologies often prioritize the extraction of emotionally legible signals, despite weak epistemological foundations, and risk medicalizing affective variance [84, 162]. Such concerns are especially acute for those with conditions marked by atypical affective expression and regulation, who may be subject to mental illness predictions generated by models that circumvent personal disclosure decisions [44, 66].

The literature reviewed here suggests that emotional privacy judgments may vary by socio-demographic characteristics—including education, race/ethnicity, gender, and mental health status—each of which can shape individuals' exposure to risk and capacity for control in digital environments. Our study builds on this body of work by empirically examining the relationship between socio-demographics and emotional privacy judgments concerning emotion inferences in workplace and healthcare contexts. We do so using two U.S. samples: a nationally representative cohort stratified by sex, age, and race, and a minoritized cohort composed of people of color, gender minorities, and individuals with lived experience of mental illness. Further details on recruitment and sampling are provided in Section 3.3.

2.3 Individual Privacy Beliefs and Emotional Sensitivity Judgments

Beliefs about privacy can impact how individuals perceive and evaluate technologies, including general privacy concerns, perceived sensitivity of information sensitivity, risk perceptions, and levels of institutional trust [103, 118, 152, 180].

2.3.1 General Privacy Concerns. Many studies measuring privacy rely on the concept of privacy concern, in part due to enduring disparities in how privacy is defined and conceptualized [28, 128]. While early scales captured privacy concern as a generalized construct [103, 157, 174, 183], a growing body of work shows these concerns are highly sensitive to context [1, 4, 120]. Consequently, general privacy concerns have limited utility in explaining context-specific privacy preferences and outcomes [111]. Still, general concern measures like the Internet User's Information Privacy Concerns (IUIPC) scale remain widely used—both as controls in privacy perception studies [103] and as predictors of related constructs such as privacy decision-making [88] and expectations [110].

2.3.2 Perceived Risk. Closely related to privacy concern is perceived privacy risk [21, 92]. While sometimes conceptualized as a global construct [63, 156], perceived risk can also be framed in terms of specific potential harms [21]. It can affect privacy judgments on an affective level: when a technology or data practice is viewed positively, it tends to be perceived as less risky and more beneficial, reducing privacy concerns overall [55, 133]. Risk perception is also a known mediator in privacy judgments. For example, in healthcare settings, Alraja et al. found that attitudes toward emerging technologies were shaped by perceptions of privacy, security, and trust, mediated through individual risk assessments [7]. This underscores the value of accounting for perceived risk in models of privacy perception.

2.3.3 Trust. Privacy and institutional trust are mutually reinforcing constructs [107, 180]. Trust in a specific institution can influence both general and context-specific privacy judgments [107], often by reducing the perceived risk of information misuse [146, 186]. At the same time, individuals' baseline privacy dispositions may precede and shape their levels of institutional trust [88, 107]. In both workplace and healthcare settings, trust plays a key role. Tolsdorf et al. found that workers' privacy perceptions in digitized workplaces were shaped by trust in their employers' handling of personal data [168], while Shen et al. showed that patients' willingness to share health information was similarly influenced by trust in healthcare organizations [153]. These findings highlight the importance of modeling institutional trust when evaluating privacy perceptions in contextually sensitive domains.

2.3.4 Data Sensitivity. Data sensitivity is best understood not as a static property, but as a belief that varies by individual traits and situational context [105, 118]. Sensitivity is closely linked to privacy risk: more sensitive data is seen as riskier and requiring stronger protections [18, 103]. Prior work suggests that people perceive emotional information [8, 142] and related data such as mental health status [175] as highly sensitive, particularly in commercial or surveillance contexts [42]. Importantly, people often underestimate the sensitivity of data *inputs*—like sensor data—while expressing strong concern about the *inferences* drawn from them. In Lee et al.'s study on mobile affective computing, participants viewed raw sensor data as relatively non-sensitive and often failed to recognize how it could be processed to reveal emotional or psychological traits. When made aware that such data could be used to reveal personal traits, however, participants expressed greater concern [95]. These findings suggest that emotional privacy judgments may be more accurately captured when people are explicitly informed about how inferences are generated—i.e., from which inputs, and for what purposes. We expand on these implications for vignette design in Section 3.2.

Our study examines how emotional privacy judgments are shaped by individual privacy beliefs. By analyzing workers' and patients' evaluations of emotional information flows in workplace and healthcare contexts, we offer insight into how individual beliefs interact with contextual features and socio-demographic characteristics. These findings inform the design of policies and systems that more closely align with the privacy judgments and needs of diverse people and groups.

3 METHODS

To investigate how workers and patients evaluate the privacy of emotion AI technologies, we designed a factorial vignette survey. Eliciting privacy perceptions can be methodologically challenging, particularly when investigating complex socio-technical systems [81, 124, 129, 185]. Factorial vignette designs offer a robust method for uncovering how privacy judgments vary across contextual and individual factors [108].

Participants rated their level of comfort with a series of vignettes in which employers and healthcare providers used data already collected about them to automatically infer their emotions. The vignettes systematically varied contextual features, while a post-test gathered socio-demographic information and privacy beliefs. This design enabled us to examine whether, and to what extent, emotional privacy judgments vary as a function of contextual, socio-demographic, and individual privacy belief factors.

The sections that follow describe the theoretical frameworks guiding the survey design, the vignette and post-test instruments, recruitment and sampling procedures, and data analysis strategy. We conclude with a reflection on limitations and opportunities for future research.

3.1 Contextual Integrity and Privacy Vulnerability as Theoretical Frameworks to Measure Emotional Privacy Judgments

Our study design draws on two theoretical frameworks. First, we operationalize *contextual integrity* using factorial vignettes—a method for measuring privacy perceptions under the theory, pioneered by Martin [108]. This approach presents scenarios structured by contextual integrity's conceptual parameters to elicit privacy judgments about information flows, enabling researchers to present scenarios where multiple variables vary systematically, allowing for the study of their combined and relative effects on privacy judgments [21, 109]. Second, we extend this methodological approach by incorporating McDonald and Forte's concept of *privacy vulnerability*, which emphasizes the unique privacy risks faced by vulnerable and minoritized groups when privacy norms operate unevenly or unjustly [114].

3.1.1 Contextual Integrity. According to contextual integrity, privacy violations occur when information flows transgress the norms of information flow specific to a given context. Establishing these norms requires specifying five parameters: (1) information type (about what); (2) subject (about whom); (3) sender (by whom); 4) recipient (to whom); and (5) transmission principles (under what conditions the information flows) [108]. These parameters operate interdependently to "predict a complex dependency between privacy judgments on the one hand, and the values for all five parameters on the other" [109]. Our study modeled this interdependency to investigate workers' and patients' emotional privacy judgments by fixing the contextual parameters that govern norms surrounding emotional information flows as follows in Table 1.

Contextual Parameter	Emotional Privacy Norms
Information Type	emotional state (e.g., emotions, moods, including but not limited to stress, anxiety, depression, boredom, calmness, fear, fatigue, attentiveness, happiness, sadness, disgust, surprise, and anger)
Subject*	employees / patients
Sender	automated systems that infer emotions and other affective phenomena by processing diverse data input modalities (i.e., emotion AI systems and related practices)
Recipient*	employer / healthcare provider
Transmission Principles	• Recipient retains subject's emotional information indefinitely, as allowed by law • Recipient will not share subjects' emotional information unless otherwise noted • Subject consented to monitoring by recipient

Table 1. Emotional Privacy Norms Specified in Vignettes, adapted from Martin and Nissenbaum (2016) [108]. **Factorial Vignette Condition.*

3.1.2 Privacy Vulnerabilities. While contextual integrity is a foundational theory in privacy research, McDonald and Forte argue that it often overlooks how privacy norms can function unevenly—benefiting privileged groups while disadvantaging vulnerable or minoritized groups. Drawing upon critical theories that expose how norms themselves can perpetuate exclusion and oppression [40, 43, 60, 99, 99, 113],they propose that HCI research move "beyond norms" to center *privacy vulnerability* as both an analytic and normative lens. This perspective recognizes how individuals' identities and social positions shape their privacy risks, which may not be reflected in dominant norms, and seeks to advance a socially just understanding of privacy that accounts for all [114].

We aligned our study with this perspective in two ways. First, our study design accounted for socio-demographic differences by quantifying the relative influence of education, race/ethnicity, gender, and mental health status on emotional privacy judgments measured using contextual integrity theory. Second, we conducted the study across two samples: a U.S. nationally representative cohort by race, sex, and age (n=300), and a cohort oversampling participants by minoritized identity statuses—race/ethnicity, gender, and mental health status (n=385).

Analyzing these groups separately allowed us to identify comparative patterns that could otherwise be obscured in pooled analyses. This methodological choice was not intended to monolithize minoritized perspectives, but rather to empirically examine McDonald and Forte's theoretical proposition: that privacy norms may obscure, reinforce, and therefore systematically disadvantage certain groups based on intersecting vulnerabilities. By disaggregating socio-demographic factors and comparing privacy judgments across dominant and minoritized cohorts, we use contextual integrity not as a static normative framework but as an empirical tool to reveal how privacy expectations may differ across social identities. This approach supports, rather than replaces, contextual integrity's foundational principles while advancing McDonald and Forte's call to empirically surface privacy vulnerabilities as both analytic and normative concerns. While our study focused on education, race/ethnicity, gender, and mental health status, we recognize that other minoritized statuses (e.g., disability, assistive technology use) are also relevant and warrant future empirical attention.

Normatively, we conceptualize vulnerability as referring to groups requiring additional protections or safeguards beyond those conventionally provided, consistent with medical and research ethics standards [11, 58]. Empirically, we define vulnerability as encompassing groups known to face significant disparities and unmet needs (e.g., risk factors, access, outcomes) in labor and health domains—including the economically disadvantaged by education; racial, gender, and ethnic minorities; and individuals with chronic health conditions including mental illness [46, 165]. Through this lens, our study centers *privacy vulnerabilities* by: (1) incorporating socio-demographic factors known to shape privacy judgments (education, race/ethnicity, gender, mental health status), and (2) adopting a dual-sampling strategy to comparatively assess emotional privacy judgments between socially dominant perspectives (i.e., U.S. representative sample) and minoritized groups known to be disproportionately surveilled, more vulnerable to privacy harms, or otherwise possessing distinct privacy needs. The literature supporting this approach is reviewed in Section 2.2.

By integrating *contextual integrity's* normative parameters with a theoretically grounded and empirically informed *privacy vulnerabilities* lens, our study investigates both how emotional information flows are evaluated and how privacy judgments vary across social positions with privacy experiences, expectations, and needs.

3.2 Survey Design

Privacy research often confronts the so-called *privacy paradox*: individuals report privacy concerns, yet their behaviors suggest otherwise [160]. One explanation for this paradox lies in how privacy is measured. Many studies fail to specify or account for the variables that shape privacy judgments, limiting the validity and interpretability of findings [108]. Other contributing factors include individuals' limited awareness of how their data is collected, repurposed, and the potential consequences of those flows [2, 97, 106]. Measurement decisions also carry broader societal implications. Public policy often relies on the conceptualizations and findings of privacy research to inform regulation [92, 108]. It is therefore essential that research designs attend to the factors that influence privacy perceptions and norms when conceptualizing, operationalizing, and measuring privacy [108].

3.2.1 Factorial Vignettes. Conventional privacy research often overlooks the contextual and individual variables that shape privacy expectations, the perception of privacy violations, and for whom those violations are most salient [108, 114]. Grounded in contextual integrity theory (Section 3.1.1), our factorial vignette design systematically incorporated these variables to investigate workers' and patients' emotional privacy judgments.

Vignette Structure. Each vignette described a scenario in which an employer or healthcare provider used data already collected about the participant to automatically infer emotions. To ensure clarity and standardization, we fixed the contextual parameters concerning consent, data retention, and sharing practices (Table 1) and provided participants with the following reference statement at the start of each vignette set:

"Emotional state" refers to your emotions and moods, including but not limited to stress, anxiety, depression, boredom, calmness, fear, fatigue, attentiveness, happiness, sadness, disgust, surprise, and/or anger. Unless otherwise noted, assume that:

(1) your employer/healthcare provider will not share your information;
(2) your information is retained indefinitely, as allowed by law;
(3) you have consented to this monitoring through a consent form.

Participants were instructed to consider their willingness to be the subject of the described technology, taking into account the type of data, its intended use, and the social context.

Within-subjects Experimental Design. The vignette design followed a 2 (Context: workplace, healthcare) x 2 (Data Input: speech/text, image/video) x 14 (Purpose) within-subjects design. All participants responded to all 56 scenarios. Vignettes were split into two sets by context. To avoid ordering effects, we randomized vignette presentation order across the three nested dimensions: (1) context, (2) data input, and (3) purpose.

Dependent Variable: Comfort. For each vignette, participants rated their comfort using a Visual Analog Scale (VAS) ranging from 0 ("very uncomfortable") to 100 = ("very comfortable"). The VAS permitted 1-unit increments, treating comfort as a continuous measure and allowing participants to respond in line with mental models of subjective experience rather than ordinal categories. VAS is widely recommended for measuring subjective phenomena due to its metacognitive sensitivity and ability to capture fine-grained judgments [61, 72, 138], while avoiding common limitations of ordinal (e.g., Likert-type) scales such as clustering and data loss [6, 22, 34, 36, 80, 170].

Although no consensus exists on the optimal dependent variable for privacy perceptions research [81, 124, 129, 185], the appropriate measure depends on the construct of interest. Studies focused on *behavioral privacy* often use willingness-to-use (e.g., [21, 82]). However, such constructs are less suitable for *normative privacy judgments*, especially in power-imbalanced settings like employment and healthcare, where choice constraints and institutional pressures may shape expressed willingness and risk obscuring underlying privacy concerns—a dynamic consistent with the bounded rationality and malleability of privacy perceptions identified by Acquisti et al. [2].

Studies eliciting normative privacy judgments commonly use either participants' comfort levels (e.g., [101, 132]) or judgments of acceptability (e.g., [109]). We selected comfort because acceptability can be shaped by adaptive preferences or resignation to constrained choices, particularly among workers and patients with limited agency over emotion inference technologies. Comfort also correlates strongly with perceived privacy risk [21]. Although Bhatia and Breaux operationalized perceived privacy risk through willingness-to-share measures, their factorial vignette studies treated these ratings as reflecting both behavioral intent and normative judgments of risk acceptability. Their finding that discomfort ratings explained up to 79% of the variance in perceived privacy risk supports the use of comfort as a valid single-item measure of privacy judgments in scenario-based designs, offering a practical and efficient proxy where more complex outcome measures may be prohibitive. Finally, selecting comfort aligns with contextual integrity's emphasis on individuals' intuitive judgments of normative appropriateness relative to contextual norms [125]. While we regard comfort as the most suitable dependent variable for this study's focus and design, future work could explore alternative constructs or multi-item measures to assess variations in emotional privacy judgments across contexts, identity characteristics, and measurement approaches.

Vignette Prompt. We asked participants to rate their level of comfort with each scenario using the following text, adapted to context. Variable levels are summarized in Table 2.

As a $C1, rate your comfort (0 = very uncomfortable to 100 = very comfortable) with your $C2 using a computer program to automatically detect your emotional states using records of $I collected from your daily activities and device use, for the purpose of $P.

Vignette Variable	Levels
Context ($C)	(1) employee* ($C1), employer** ($C2), work performance ($C3)
	(2) patient* ($C1), healthcare provider** ($C2), overall health ($C3)
Data Input ($I)	(1) what you say (either verbally or written/typed) and how you say it (e.g., speed, tone)
	(2) images or video of what you look like, based on your facial expressions
Purpose ($P)	(1) giving ($C2) data-driven insights into ($C1)'s well-being
	(2) sharing that information with academic researchers
	(3) diagnosing mental illness in ($C1) earlier than otherwise possible
	(4) diagnosing neurological disorders (e.g., dementia, ADHD) in ($C1) earlier than otherwise possible
	(5) avoiding subjectivity in other methods ($C2) may use to learn about your emotional state (e.g., surveys, observations)
	(6) inferring mental health state of ($C1) individually
	(7) inferring mental health at the group level only
	(8) identifying ($C1) needing mental health support to better plan organizational mental health resources
	(9) inferring ($C1) at risk of harming others
	(10) inferring ($C1) at risk of self-harm
	(11) developing intelligent computer therapy programs for ($C1)
	(12) detecting moments ($C1) may need emotional support and responding to help
	(13) alerting ($C2) when ($C1) may need support
	(14) assessing ($C3) of individual ($C1)

Table 2. Vignette Variables and Levels by Contextual Factor.
**Contextual Integrity Parameter: Data Subject;*
***Contextual Integrity Parameter: Data Recipient*

We used the phrase "computer program to automatically detect emotional state" to promote neutrality and comprehension. This phrasing avoided technical jargon (e.g., "emotion AI"), stigmatizing proxies (e.g., "mental health state"), or unfamiliar terms (e.g., "affective state") that could bias or confuse participants.

Participants saw only the text relevant to the given context ("*as an employee...*" or "*as a patient...*") crossed with the assigned data input, followed by a separate VAS slider for each of the 14 purposes. This design minimized cognitive load and improved response efficiency. Although participants rated 56 vignettes (28 per context), the consistency of the purposes across data input conditions allowed participants to develop familiarity with the response format and proceed quickly. Figure 1 shows an example vignette from the employment context.

Purpose Selection. The 14 purposes included in this study were selected through a two-stage process to ensure relevance, validity, and comparability across contexts. First, we identified common and emerging uses of emotion AI documented in industry practice through patent analyses of workplace [27] and healthcare [85] applications. Second, we cross-referenced these industry uses with scholarly literature describing potentially beneficial applications of emotion AI across both settings. This literature emphasizes purposes such as providing general well-being insights and more

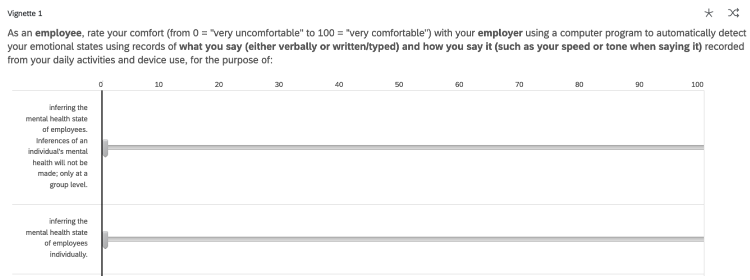

Fig. 1. Presentation of Vignettes for the Employment Context, Partial Example

specific mental health inferences at both individual and group levels [33, 90]; detecting or preventing self-harm or harm to others [44, 142, 148, 166]; and enabling early detection of mental and neurological illness with the goal of improving mental health support, safety monitoring, and research [44, 66, 90, 115, 142, 166].

While comprehensive, this set is not intended to be exhaustive. Rather, the purposes represent a theoretically and empirically grounded sample of common and proposed emotion AI uses relevant to workplace and healthcare contexts. In our analysis, purpose is modeled as a fixed effect, reflecting these specific uses, rather than as a random sample intended to generalize to all conceivable purposes. This design choice supports the validity of our findings while acknowledging that other uses warrant future empirical attention.

3.2.2 Open-ended questions. After completing each of the two vignette sets (employment and healthcare contexts), participants answered four open-ended questions:

(1) In what ways, if any, do you think these systems could benefit you? Please describe and provide examples and as much detail as you are comfortable with.
(2) In what ways, if any, do you think these systems could harm you or have other undesired impacts on you? Please describe and provide examples and as much detail as you are comfortable with.
(3) What other concerns, if any, do you have about these systems? Please describe and provide examples and as much detail as you are comfortable with.
(4) In what ways, if at all, do aspects of who you are (for example, your race/ethnicity, gender, sexuality, employment status, class, education, mental health conditions, physical health conditions, or any other features of your identity) shape your responses to the use of computer programs to infer your emotional states?

The qualitative data gathered through these questions provided rich insights into participants' perceived benefits, risks, and personal contexts influencing their privacy judgments. Due to the volume and depth of the qualitative responses, this analysis was beyond the scope of the current quantitative study focused on measuring emotional privacy judgments, and was instead addressed in two separate, peer-reviewed publications [41, 143]. Where relevant, we briefly reference key qualitative themes in the present manuscript to complement interpretation of the quantitative results. Full methodological details, coding strategies, and findings from the qualitative analyses are available in the cited publications.

3.2.3 Post-test. After completing the vignette ratings and open-ended questions, participants responded to a post-test that gathered additional information about individual characteristics. Following best practices for inclusive survey data collection [31, 56, 65, 161], the post-test collected socio-demographic information, including race/ethnicity, gender, age, subjective socio-economic status, mental health status, employment status, and educational attainment.

The post-test also assessed individual privacy beliefs. Participants responded to items measuring general information privacy concerns, perceived risks of employer and healthcare provider access to sensitive personal information, institutional trust in those entities, and the perceived sensitivity of emotional information relative to other commonly recognized sensitive data types [38, 134]. These items adapted the Internet Users' Information Privacy Concerns (IUIPC) scale [103] to our specific contexts of employment and healthcare. Full item wording for the socio-demographic and privacy belief measures appears in Appendices B and C. Participants used the same Visual Analog Scale (VAS) ranging from 1 to 100 to report privacy beliefs, maintaining consistency with the vignette ratings. To avoid potential priming effects, we administered the post-test only after participants completed all vignette responses.

These measures allowed us to analyze whether, and how, socio-demographic and privacy belief factors shaped emotional privacy judgments alongside the contextual integrity parameters and additional contextual factors varied in the vignettes.

3.2.4 Pilot Study. To ensure the survey consistently measured the intended constructs, we conducted a pilot study (n=25) in which participants completed the survey vignettes and provided feedback on any confusing elements. Analysis of the pilot data indicated that no substantive design changes were necessary. Participants' responses confirmed that their comfort ratings reflected perceptions of employer or healthcare provider use of computational emotion inferences specifically, rather than general monitoring—supporting the survey's construct validity.

The pilot also assessed potential participant fatigue. We included attention check questions and monitored completion time. While factorial vignette designs often entail a learning curve due to their novelty rather than respondent fatigue [112]; participants became familiar with the vignette format quickly. Despite evaluating 56 vignettes, the average completion time was 24 minutes, and only two participants failed the attention check. These results indicated the survey length was appropriate for the study's objectives.

3.3 Recruitment and Data Collection

3.3.1 Sampling. We collected two samples to assess emotional privacy judgments: (1) a U.S. nationally representative sample by age, sex, and race (n=300), and (2) a sample oversampling individuals with one or more minoritized identities (person of color, minority gender, and/or mental illness status; n=385). As described in Section 3.1.1, this sampling strategy allowed us to investigate how privacy judgments vary both within and between socially dominant and minoritized perspectives.

3.3.2 Recruitment. Participants were recruited via Prolific, using pre-screening criteria for age, sex, race, minoritized identities, and other relevant characteristics. The nationally representative participant group was recruited in October 2021 using Prolific's automatic balancing feature. The minoritized participant group was recruited between December 2021 and February 2022 using targeted pre-screening. Participants completed the survey through Qualtrics and were compensated $3.80, following Prolific's recommended rate. We note that some under-represented gender and ethnic minority groups could not be analyzed separately due to small sample sizes. Summary statistics are reported in Table 3.

3.3.3 Ethical Oversight. Our institution's IRB determined that this study qualified for exemption from oversight under 45 CFR 46.104(d)(2)(i), which applies to survey procedures where information is recorded such that subjects cannot readily be identified, directly or indirectly [173]. Data were collected anonymously via the Prolific platform, which

Factor	Level	Representative Sample	Minoritized Sample
Race/Ethnicity	Additional ethnicities	11	26
	Asian	26	47
	Black	51	104
	Latine	15	42
	White	197	194
Gender	Trans and/or non-binary	6	44
	Woman	148	232
	Man	146	139
Mental Health Status	Under treatment for 1+ mental illness	67	115
	Untreated/resolved mental illness	50	101
	No mental illness	183	140
	Did not report	0	57
Age Group	18–27	55	170
	28–37	55	120
	38–47	49	42
	48–57	52	30
	58+	89	36
	Did not report	0	15
Education	Bachelor's degree or higher	170	167
	No bachelor's degree	130	190
	Did not report	0	56

Table 3. Descriptive Sample Statistics by Socio-demographic Level

compensated participants directly, eliminating the need for researchers to collect linkable personal information. The study was determined to involve no more than minimal risk to participants, consistent with federal research ethics standards.

Exemption from oversight does not preclude ethical responsibility. The research team followed best practices to protect participant privacy, data security, and dignity, including obtaining informed consent, ensuring anonymity in survey responses, minimizing participant burden, and reviewing pilot study results for potential ethical concerns. The pilot study identified no design or content issues, and participants' open-ended responses indicated high engagement and willingness to reflect on the study topics. Although Prolific assigns participant IDs, these were not published or linked to study results, and data access was restricted to the research team.

3.4 Data Analysis

3.4.1 Pre-processing. We prepared the dataset by removing 49 respondents who did not complete both vignette sets, 13 who did not provide demographic information, one who failed the attention check, one low-quality response (identical answers across all vignettes without justification in open-ended responses), and 12 duplicate submissions. When duplicate entries occurred, we retained the most complete submission or, when both were complete, the first submission.

Missing data were minimal (60 missing responses), occurring only in post-test variables. No missing data occurred in the vignette ratings or socio-demographic variables used in the analysis. Given this low rate and the non-systematic nature of missingness, we applied mode imputation for the categorical privacy belief variables using the *mice* package in R, a common method in social science research for handling sparse missing data [49, 171].

To ensure adequate statistical power and interpretability, we condensed the socio-demographic categories collected in the post-test (Appendix B). Final groupings are detailed in Table 3. Because of inconsistencies between the race/ethnicity values reported in our pre-screener and those recorded by Prolific (Section 3.3.2), participants reporting mixed or multiple race/ethnicities were classified according to either their non-white race/ethnicity or primary ethnicity to maintain data integrity. For instance, participants identifying as both white and Latine in the pre-screener were inconsistently categorized by Prolific (e.g., "white," "mixed," or "other"). In alignment with scholarship critiquing the treatment of Latine identities in U.S. race reporting [141], we coded these participants as Latine. Participants reporting multiple non-white identities were assigned to their primary race/ethnicity as reported in Prolific, as these cases exhibited fewer inconsistencies.

3.4.2 Factors. For both the representative and minoritized samples, we regressed contextual, socio-demographic, and individual privacy belief variables on participants' reported comfort ratings for each vignette. Table 4 summarizes the factors included in the analysis.

Factor Type	Variables
Contextual	• Context • Data Input • Purpose
Socio-demographic	• Race/Ethnicity • Gender • Mental Health Status • Educational Attainment
Individual Privacy Beliefs	• General Privacy Concerns • Trust in Employer/Healthcare Provider Handling of Sensitive Information • Perceived Sensitivity of Emotional Information in Employment/Healthcare

Table 4. Analysis Factors and Variables

For individual privacy beliefs reported in the post-test, we averaged participants' responses (0–100) across each belief construct: general privacy concerns, trust in employer/healthcare provider data practices, and perceived sensitivity of emotional information. Responses to negatively worded items were reverse-coded where necessary to ensure conceptual consistency. Categorical predictors (e.g., purpose, gender, race/ethnicity) were treated as factors and internally contrast coded (default treatment coding) by the *lme4* package. Continuous predictors (e.g., privacy belief variables) retained their original scale metrics and were modeled as continuous fixed effects.

3.4.3 Mixed Effect Modeling. To examine how contextual, socio-demographic, and privacy belief factors influenced emotional privacy judgments, we used multivariable linear mixed-effects models fitted in R using the *lme4* package.

Because each participant rated multiple vignettes, we specified participants as random effects to account for within-subject correlation, limit biased covariance estimates, and avoid violating independence assumptions [54, 64]. Fixed effects included the variables listed in Table 4. We fit four multi-variable linear mixed-effects models, one for each combination of employment/healthcare vignette sets and representative/minoritized samples. This structure was necessary because individual privacy belief variables differed between contexts and because separate samples allowed for comparison across social groups.

Model selection followed standard multi-level modeling procedures. We compared candidate models using likelihood ratio tests (LRTs), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) [17, 136]. Variables were retained as fixed effects if they significantly improved model fit in at least one of the four datasets. We used maximum likelihood estimation (ML) to enable comparisons between models with different fixed effects structures [136]. Final models included the same set of fixed effects across samples to facilitate direct comparison.

Socio-demographic reference categories reflected the dominant social group in each category: white, male, age 58+, no mental illness, and bachelor's degree or higher. For purpose, we used "giving employers/healthcare providers data-driven understanding into employee/patient well-being" as the reference, reflecting typical workplace and healthcare justifications for emotion inference technologies [100, 123].

We tested statistical significance using t-tests with Satterthwaite's approximation [149]. We verified model fitness by comparing each final model to a null model and calculated intra-class correlations (ICC) to assess variance explained by the random effects structure [79]. The ICCs (.72 and .67) indicated good reliability [154]. Residuals were plotted to confirm normality assumptions, which mixed models are robust against even when variables are not normally distributed [12, 78, 178].

We selected random slope models, allowing factor effects to vary across participants. All fixed effects are listed in Table 4. Variables excluded from the final models included age group, subjective socio-economic status, and employment status, which did not improve model fit. Participants' privacy risk beliefs were excluded due to multicollinearity with trust beliefs. Although general privacy concerns did not significantly predict outcomes in any of the four models, we retained this variable given its theoretical relevance to privacy perceptions.

To compare factor effects between samples, we used Z-tests. A positive Z-score indicated stronger effects in the representative sample; a negative Z-score indicated stronger effects in the minoritized sample. Significant differences were identified at $|Z| > 1.96$ ($p < .05$). Even when effects were not statistically significant, variation between groups revealed meaningful patterns in how different factors shaped privacy judgments.

3.5 Reflections, Limitations, and Opportunities

3.5.1 Survey Responses. Our design elicited workers' and patients' self-reported comfort with being subject to various applications of automatic emotion inferences as a measure of their emotional privacy judgments. We framed vignettes as neutrally as possible, avoiding references to potential harms. However, some purposes (e.g., enhancing safety or mental health support) may have implied benefits, which could have influenced judgments [21]. Future work could test framing effects more explicitly.

Our use of a continuous Visual Analog Scale (VAS) for the dependent variable reduced common limitations of ordinal scales, such as data loss and clustering (see Section 3.2.1). While standard limitations of self-reported data apply, factorial vignette designs mitigate respondent bias by varying factors across scenarios, making it difficult for participants to systematically adjust responses [108].

3.5.2 Model Variables and Missing Factors. We recognize that privacy judgments are shaped by a wide range of contextual and individual factors. Our models focused on contextual integrity parameters, socio-demographic identities, and privacy beliefs most relevant to our research questions. Our vignettes specified consent and data handling parameters consistent with typical workplace and healthcare data practices. However, real-world implementations may involve organizational and institutional cultures, different consent dynamics, data sharing policies, or types of emotional information, which could affect privacy judgments in addition to individual variables such as privacy awareness or technological literacy. Such factors were beyond the scope of this study but merit future investigation.

3.5.3 Generalizability and Sample Limitations. While our U.S. representative sample followed standard demographic balancing procedures and our minoritized sample intentionally centered marginalized perspectives, neither fully captures the diversity of experiences within these participant groups.

Participants were recruited from Prolific—click workers who are often over-represented in research and whose privacy perceptions may differ from the broader population. Nonetheless, recent scholarship indicates that Prolific samples are generally representative in studies of privacy perceptions [164], supporting the validity of findings drawn from our U.S. representative sample.

Finally, our decision to combine data input types and to examine emotion inferences without specifying emotion categories facilitated a manageable vignette design and minimized participant fatigue. However, these choices necessarily limit the granularity of our findings. Future research should explore how privacy judgments vary across more specific data modalities and emotion types.

3.5.4 Statistical Considerations. Our mixed-effects models balanced theoretical relevance with statistical rigor, accounting for individual variability and interdependent predictors. As expected in models incorporating multiple variables and random effects, some factors showed non-significant associations [24, 136]. We interpret these conservatively and report confidence intervals to avoid dichotomous significance testing [24, 136]. Where appropriate, we discuss notable patterns that may have theoretical significance [24, 102].

4 FINDINGS

Our study systematically dissects the complex interplay of factors influencing emotional privacy judgments toward technologies that infer and interact with human emotion in workplace and healthcare settings. Using mixed-effects modeling, we examine how contextual, socio-demographic, and individual privacy belief factors differentially influence workers' and patients' comfort levels. Our findings synthesize insights crucial to privacy theory, human-computer interaction, and technology policy, enhancing our understanding of emotional privacy amid growing AI-driven practices.

Recognizing that privacy perceptions vary across contexts and between dominant (U.S. representative) and minoritized groups [114, 124], we highlight these variations to underscore the multi-dimensional nature of emotional privacy judgments. Our rigorous methodological framework (see Sections 3.1 and 3.4.3) enables meaningful comparisons in emotional privacy judgments and identification of significant trends and differences.

Regression results, summarized in Tables 5 (employment) and 6 (healthcare), present coefficients, standard errors, and statistical significance across key variables.

Regression Results for Employment Context

	Representative (n=300)	Minoritized (n=385)	Z-Test (Sample Comparison)
(Intercept)	36.64 (6.44)***	34.24 (6.08)***	0.27
Contextual Factors			
Data Input *(baseline: image/video)*			
speech/text	2.69 (0.35)***	4.25 (0.34)***	**-3.21**
Purpose *(baseline: (1) data-driven well-being insights)*			
(2) academic research	4.18 (0.93)***	1.26 (0.89)	**2.28**
(3) early diagnosis - mental illness	−1.32 (0.93)	−2.49 (0.89)**	0.91
(4) early diagnosis - neurological disorder	0.55 (0.93)	−1.70 (0.89)$^{\cdot}$	1.75
(5) avoid human subjectivity	0.57 (0.93)	−1.39 (0.89)	1.52
(6) individual level mental health inference	−3.48 (0.93)***	−3.70 (0.89)***	0.17
(7) group level mental health inference	2.63 (0.93)**	2.16 (0.89)*	0.36
(8) identify individuals in need of support	2.15 (0.93)*	3.78 (0.89)***	−1.27
(9) infer risk of harm toward others	6.39 (0.93)***	7.03 (0.89)***	−0.50
(10) infer risk of self-harm	3.20 (0.93)***	2.60 (0.89)**	0.47
(11) develop auto. intervention - therapy	1.68 (0.93)$^{\cdot}$	1.92 (0.87)*	−0.19
(12) receive auto. intervention - acute support	1.66 (0.93)$^{\cdot}$	1.85 (0.89)*	−0.14
(13) alert employer when in need of support	0.28 (0.93)	−0.05 (0.89)	0.26
(14) assess overall performance	−0.88 (0.93)	−2.55 (0.89)**	1.30
Socio-demographic Factors			
Race/Ethnicity *(baseline: white)*			
Asian	−3.15 (4.31)	−8.05 (3.55)*	0.88
Black	5.64 (3.19)$^{\cdot}$	7.38 (2.79)**	−0.41
Latine	8.27 (5.44)	4.45 (3.84)	0.57
additional races/ethnicities	6.78 (6.27)	3.53 (4.69)	0.41
Gender *(baseline: male)*			

	Representative (n=300)	Minoritized (n=385)	Z-Test (Sample Comparison)
trans and/or non-binary	−0.63 (8.71)	−4.63 (4.08)	0.42
woman	−2.99 (2.41)	−0.06 (2.45)	−0.85
Mental Health *(baseline: no mental illness)*			
under treatment for 1 or more mental illnesses	6.47 (3.13)*	−0.79 (3.06)	1.66
used to have/not treated for mental illness	−3.67 (3.36)	2.17 (2.97)	−1.30
Education *(baseline: Bachelor's or higher)*			
no Bachelor's degree or less	0.14 (2.46)	6.16 (2.37)**	−1.76
Individual Privacy Beliefs			
general privacy concerns	−0.04 (0.07)	−0.07 (0.07)	0.34
sensitivity toward emotion data in employment	−0.30 (0.05)***	−0.25 (0.05)***	−0.71
trust beliefs re: employer sensitive info	0.54 (0.05)***	0.40 (0.05)***	**2.09**
Akaike Information Criterion (AIC)	71657.54	93685.87	
Bayesian Information Criterion (BIC)	71861.58	93911.72	
Log Likelihood	−35799.77	−46811.94	
Number of observations	8400	10780	
Number of groups: Individual participants	300	385	
Var: Individual participant (Intercept)	394.19	417.98	
Var: Residual	257.45	303.61	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$; **Bold** Z-scores indicate statistical significance between samples at p<0.05

Table 5. Regression results: Effect of Independent Variables on Comfort with Emotion Inferences in Employment

Regression Results for Healthcare Context

	Representative (n=300)	Minoritized (n=385)	Z-Test (Sample Comparison)
(Intercept)	28.91 (8.21)***	26.96 (6.72)***	0.18
<u>Contextual Factors</u>			
Data Input *(baseline: image/video)*			
speech/text	4.13 (0.37)***	5.35 (0.36)***	**-2.37**
Purpose *(baseline: (1) data-driven well-being insights)*			
(2) academic research	−2.61 (0.98)**	−2.43 (0.95)*	−0.13
(3) early diagnosis - mental illness	−1.98 (0.98)*	−0.55 (0.95)	−1.05
(4) early diagnosis - neurological disorder	2.19 (0.98)*	3.47 (0.95)***	−0.94
(5) avoid human subjectivity	−3.73 (0.98)***	−2.44 (0.95)**	−0.94
(6) individual level mental health inference	−5.52 (0.98)***	−4.72 (0.95)***	−0.59
(7) group level mental health inference	−4.32 (0.93)***	−4.74 (0.95)***	0.31
(8) identify individuals in need of support	−1.17 (0.98)	1.34 (0.95)	**-1.84**
(9) infer risk of harm toward others	−0.26 (0.98)	−0.56 (0.95)	0.22
(10) infer risk of self-harm	−0.27 (0.98)	−1.10 (0.95)	0.61
(11) develop auto. intervention - therapy	−7.91 (0.98)***	−7.88 (0.95)***	−0.02
(12) receive auto. intervention - acute support	−3.49 (0.98)***	−3.18 (0.95)***	−0.23
(13) alert provider when in need of support	−3.89 (0.98)***	−2.09 (0.95)*	−1.32
(14) assess overall health	−1.91 (0.98)$\cdot$	0.23 (0.95)	−1.57
<u>Socio-demographic Factors</u>			
Race/Ethnicity *(baseline: white)*			
Asian	3.33 (5.44)	−3.90 (3.92)	1.08
Black	10.65 (4.02)**	6.66 (3.05)*	0.79
Latine	4.60 (6.87)	5.47 (4.29)	−0.11
additional races/ethnicities	3.01 (7.93)	0.95 (5.18)	0.22
Gender *(baseline: male)*			

	Representative (n=300)	Minoritized (n=385)	Z-Test (Sample Comparison)
trans and/or non-binary	−6.26 (10.99)	−15.32 (4.55)***	0.76
woman	−1.52 (3.03)	−0.07 (2.71)	−0.36
Mental Health *(baseline: no mental illness)*			
under treatment for 1 or more mental illnesses	3.69 (3.94)	1.70 (3.33)	0.39
used to have/not treated for mental illness	−0.12 (4.23)	3.32 (3.31)	−0.64
Education *(baseline: Bachelor's or higher)*			
no Bachelor's degree or less	2.06 (3.06)	2.12 (2.63)	−0.01
Individual Privacy Beliefs			
general privacy concerns	−0.04 (0.08)	−0.08 (0.07)	0.35
sensitivity toward emotion data in healthcare	−0.10 (0.05)$^{\cdot}$	−0.11 (0.04)**	0.15
trust beliefs re: healthcare sensitive info	0.44 (0.06)***	0.53 (0.05)***	−1.07
Akaike Information Criterion (AIC)	72732.10	94520.25	
Bayesian Information Criterion (BIC)	72936.15	94745.93	
Log Likelihood	−36337.05	−47229.12	
Number of observations	8400	10724	
Number of groups: Individual participants	300	385	
Var: Individual participant (Intercept)	632.11	511.98	
Var: Residual	288.95	342.41	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$; **Bold** Z-scores indicate statistical significance between samples at p<0.05

Table 6. Regression results: Effect of Independent Variables on Comfort with Emotion Inferences in Healthcare

Together, these analyses provide a comprehensive overview of how contextual, socio-demographic, and privacy belief factors influence emotional privacy judgments about emotion AI in the workplace and healthcare.

Table 7 distills these results, complemented visually by the coefficient plot in Figure 8 (Appendix A).

Factor	Key Findings	Sample Differences (Rep. vs Minoritized)
Context	Lower baseline comfort in employment than healthcare. Contextual factors exert greater influence in healthcare.	Minoritized groups generally reported lower comfort across contexts.
Data Input	Speech/text preferred over image/video inputs across both contexts.	Stronger preference for speech/text in minoritized samples.
Purpose	Strong effects by purpose. Group-level inferences, harm prevention, and academic research (employment only) raised comfort. Individual assessments and mental health diagnostics reduced comfort, especially in healthcare.	Minoritized groups showed lower comfort across most purposes. Greater trust in neurological diagnostics in healthcare.
Race/Ethnicity	Black participants reported higher comfort across contexts; Asian participants lower comfort in employment.	Black participants more positive; Asian and minoritized participants more cautious, especially in employment.
Gender	No significant effects except trans/non-binary participants reporting lower comfort in healthcare (strong effect in minoritized sample).	Strong negative effect for trans/non-binary in healthcare.
Mental Health Status	Current treatment increased comfort in employment (rep. sample only). No significant effects in healthcare.	Divergence between samples; effect attenuated or reversed in minoritized group.
Education	Lower educational attainment linked to higher comfort in employment (significant in minoritized sample).	Larger positive effect in minoritized sample.
General Privacy Concerns	No significant effects.	N/A
Trust Beliefs	Higher trust increased comfort in both contexts.	Stronger effect in representative sample (employment); stronger in minoritized sample (healthcare).
Perceived Sensitivity of Emotional Data	Higher perceived sensitivity linked to lower comfort across contexts.	Effect confirmed in minoritized sample (healthcare); similar trends elsewhere.

Table 7. Summary of Key Quantitative Findings Across Factors and Sample Comparisons.

4.1 Influence of Contextual Factors on Emotional Privacy Judgments of Emotion Inferences in the Workplace and Healthcare

In examining emotional privacy judgments concerning emotion inferences, we focused on three key contextual variables: context ($C), data input ($I), and purpose ($P). Participants evaluated tailored vignettes that varied by these variables as follows:

As a $C1, rate your comfort (from 0 = "very uncomfortable" to 100 = "very comfortable) with your $C2 using a computer program to automatically detect your emotional states using records of $I recorded from your daily activities and device use, for the purpose of $P

Our analysis assesses how data input ($I) and purpose ($P) shape emotional privacy judgments within each context.

4.1.1 Context: Emotional Privacy Judgments More Susceptible to Factor Influences in Healthcare than in Employment.

Privacy perceptions differed substantially between employment and healthcare contexts (Table 8).

Context	Sample	Mean	Mean StdDev	Regression Intercept
employment	representative	32.50	32.59	36.64
employment	minoritized	32.55	32.11	34.24
healthcare	representative	49.70	32.45	28.91
healthcare	minoritized	50.02	32.54	26.96

Table 8. Summary Statistics - Mean and Estimated Comfort Levels by Context and Sample

Mean comfort was markedly lower in employment (32.50/32.55) than in healthcare (49.70/50.02). However, the regression intercepts—which control for all other variables—reveal that in healthcare, baseline comfort was even lower (28.91/26.96). This suggests that factors in our model had a greater impact on comfort levels in healthcare than in employment, where intercepts were closer to the mean comfort levels.

These differences reflect distinct power dynamics and privacy expectations. Healthcare is anchored in trust and confidentiality, particularly around mental health, where subjective emotional disclosures are central. This reliance may amplify privacy sensitivities, especially among minoritized groups who have faced inequitable care or stigma. Our related qualitative findings confirm heightened concerns about emotion AI's potential to undermine autonomy, care access, and the patient-provider relationship [143].

By contrast, employment reflects normalized surveillance and limited worker autonomy, contributing to baseline discomfort with emotion inferences. Qualitative data indicate that workers—especially from marginalized groups—view such technologies as likely to exacerbate existing privacy and power disparities [41].

These patterns underscore the contextual variability of emotional privacy judgments and validate our use of mixed-effects modeling to disentangle how contextual, socio-demographic, and belief factors shape these judgments. Lower healthcare intercepts indicate that such factors exert stronger influences in healthcare, where participants expressed overall higher comfort yet rejected most specific purposes for emotion inference—especially those that undermined autonomy or discretion. In contrast, while employment settings evoked lower baseline comfort, participants differentiated sharply between acceptable and unacceptable uses. Some purposes—such as group-level inferences or harm prevention—elicited positive responses, suggesting conditional acceptance even in surveillance-prone environments. Yet overall, emotion inferences in employment remained a source of concern, particularly given the risks of employer misuse and the potential to reinforce existing power asymmetries.

4.1.2 Data Input: Workers and Patients Favor Speech/Text Emotion Recognition Over Facial Emotion Recognition, though Emotional Privacy Judgments Remain Low with All Modalities.

We examined whether and how participants' comfort with emotion inferences varied by the type of data input to the emotion recognition algorithm. From their perspectives as employees and patients, participants rated their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable) with their employers and healthcare providers using a computer program to detect their emotional states from either (1) speech/text records—what they say (verbally or written/typed) and how they say it (e.g., speed or tone)—or (2) images/video of their facial expressions, for various purposes.

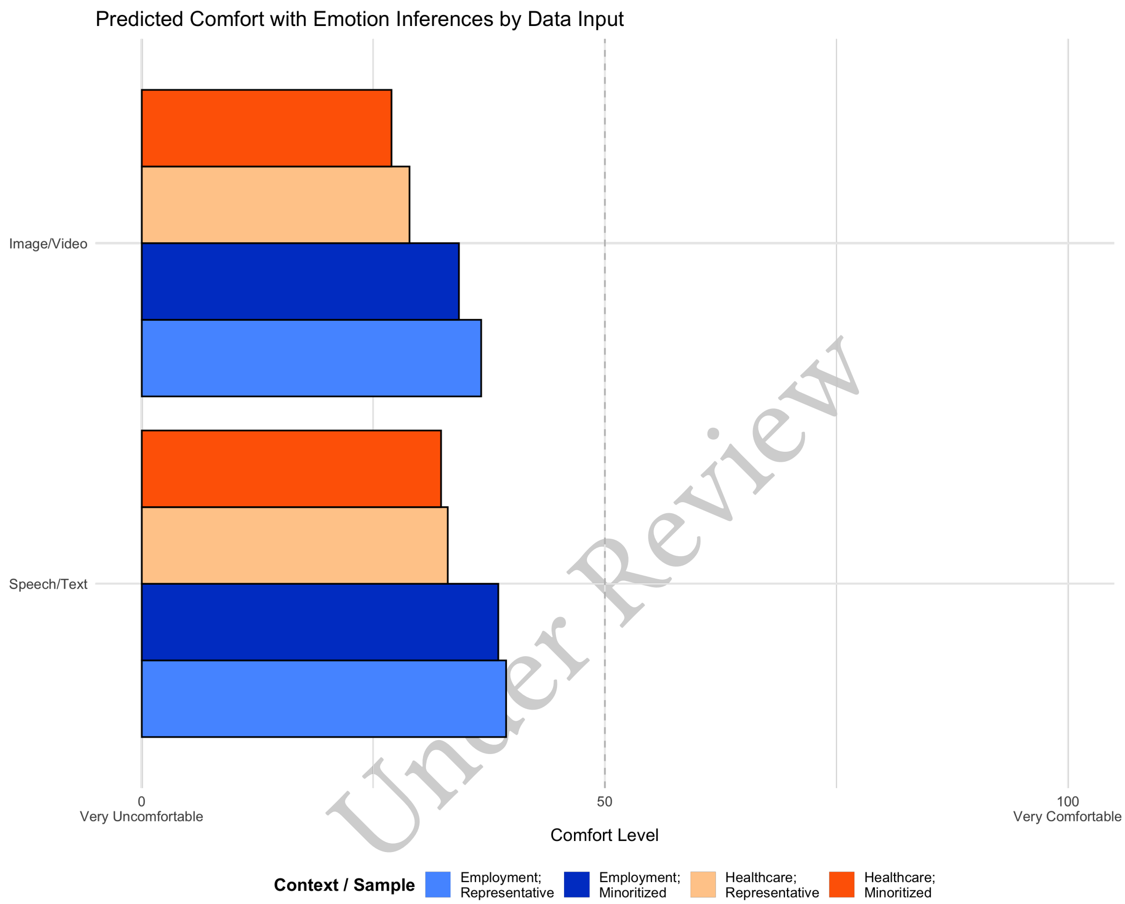

Fig. 2. Predicted Comfort Levels by Data Input Type. This figure illustrates the predicted comfort levels by combining the data type variable coefficients to each mixed-effects regression model intercept, derived by analyzing respondent comfort on a scale from 1 (very uncomfortable) to 100 (very comfortable). Bars with black borders indicate statistically significant results.

Our regression results show that both workers and patients were significantly more comfortable with speech/text-based emotion inferences than with those based on facial recognition. Compared to the baseline category of image/video records, speech/text inputs were associated with significantly higher comfort in both employment (representative: $\beta = 2.69$, $SE = 0.35$, $p < 0.001$; minoritized: $\beta = 4.25$, $SE = 0.34$, $p < 0.001$) and healthcare (representative: $\beta = 4.13$, $SE = 0.37$, $p < 0.001$; minoritized: $\beta = 5.35$, $SE = 0.36$, $p < 0.001$). This may reflect public discomfort with facial recognition technologies and their attendant accuracy and privacy concerns [191].

Notably, although speech/text inputs raised comfort relative to facial recognition, predicted comfort levels across all data inputs remained low—ranging from 32.31 to 39.33 on a 0–100 scale (Figure 2). The more pronounced positive effect of speech/text was statistically significant in both employment and healthcare, with Z-scores of -3.21 and -2.37, respectively.

These findings support a growing recognition that facial recognition technologies—including facial *emotion* recognition—are widely viewed with suspicion. They also confirm that data input is a meaningful and statistically significant contextual factor shaping emotional privacy judgments. However, this does not suggest that emotional privacy can be preserved by avoiding facial inputs alone. Even with speech/text, predicted comfort remained low.

Importantly, the effect of data input was more pronounced for participants in the minoritized sample, who consistently reported lower comfort across both contexts and all input types. This pattern underscores greater emotional privacy concerns about all forms of emotion recognition—speech, text, and facial—among people of color, people with mental illness, and/or minority genders compared to the U.S. representative cohort.

4.1.3 Purposes for Which Employers and Healthcare Providers Use Emotion Inferences Shape Emotional Privacy Judgments.

To assess the influence of purpose on emotional privacy judgments, we examined whether and how participants' comfort varied across fourteen distinct purposes for which employers and healthcare providers might use emotion inferences (Table 2). We report how participants rated their comfort (0 = "very uncomfortable" to 100 = "very comfortable) relative to a common baseline: providing data-driven insights into employee or patient well-being. For interpretive clarity, we grouped the fourteen purposes into higher-level themes (Table 9).

Our findings demonstrate that purpose significantly shapes emotional privacy judgments, with effects varying by specific purpose, context, and participant group. Generally, purposes that reinforced each context's social mission or aligned with its privacy expectations were judged more positively, while purposes that strained those expectations were judged more negatively. As prior work shows, perceived technological benefits and risks can both influence privacy perceptions [21].

To contextualize these findings, we draw on qualitative analyses of perceived benefits and risks voiced by study participants, drawn from their open-ended responses (Section 3.2.2) and published in related studies on employment [41] and healthcare [143].

Figure 3 visualizes predicted comfort levels for each purpose.

Facilitating Earlier Diagnosis of Neurological Disorders and Mental Illness. One proposed use case for emotion inferences involves facilitating earlier medical diagnosis. This application has been suggested for healthcare and, increasingly, the workplace—given the extensive time people spend at work and the rise of surveillance systems already collecting data from which emotional features might be extracted [32, 91, 135]. We examined how using emotion inferences to detect mental illnesses and neurological disorders earlier than otherwise possible influenced participants' comfort. Participants rated their comfort (0 = "very uncomfortable" to 100 = "very comfortable") with their employers or healthcare providers using emotion inferences for:

> *(3) diagnosing neurological disorders, such as dementia or ADHD, in employees/patients earlier than otherwise possible*; and
>
> *(4) diagnosing mental illness in employees/patients earlier than otherwise possible.*

Purpose Grouping	Purpose Levels
early diagnosis of mental illness and neurological disorders	(3) diagnose mental illness in ($C1) earlier than otherwise possible (4) diagnose neurological disorders (e.g., dementia or ADHD) in ($C1)
augment employee and patient assessments	(5) avoiding subjectivity in other methods your ($C2) may use to learn about your emotional state, like a survey or your ($C2)'s observation (14) assessing the ($C3) of individual ($C1)
individual and group-level mental health inferences	(6) inferring the mental health state of ($C1) individually (7) inferring the mental health state of ($C1). inferences of an individual's mental health will not be made; only at a group level
societal benefit	(2) sharing that information with academic researchers to help them learn more about mental health, as part of a research partnership (8) identifying ($C1) in need of mental health support, to better plan organizational mental health resources
harm prevention	(9) inferring whether ($C1) are at risk of harming others (10) inferring whether ($C1) are at risk of harming themselves
supportive interventions	(11) developing an intelligent computer program, such as a chat bot, that can conduct mental health therapy with ($C1), including you (12) inferring moments ($C1) may be in need of emotional support and responding with an intelligent computer program designed to help ($C1) improve their well-being, such as offering well-being tips (13) automatically alerting your ($C2) when ($C1)s may need support, including you
baseline purpose	(1) giving ($C2) data-driven insights into ($C1) well-being

Table 9. We examined the impact of 14 purposes for which employers and healthcare providers may use emotion inferences, grouped into higher level themes to aid interpretation.

Predicted comfort levels for both diagnostic purposes remained low across contexts and samples—ranging from 31.75 to 37.19 for workers and 26.41 to 31.1 for patients. As shown in Figure 3, comfort was consistently lower in healthcare than in employment, reflecting heightened privacy concerns about emotion inferences in clinical settings.

Earlier diagnosis of mental illness. Across both contexts and samples, using emotion inferences to detect *mental illness* had a negative effect on comfort compared to the baseline purpose. While both workers and patients expressed discomfort with this application, differences emerged by context and sample.

Employment context. For employment, the negative effect was statistically significant only in the minoritized sample (representative: $\beta = -1.32$, $SE = 0.93$, not significant; minoritized: $\beta = -2.49$, $SE = 0.89$, $p < 0.01$). Qualitative

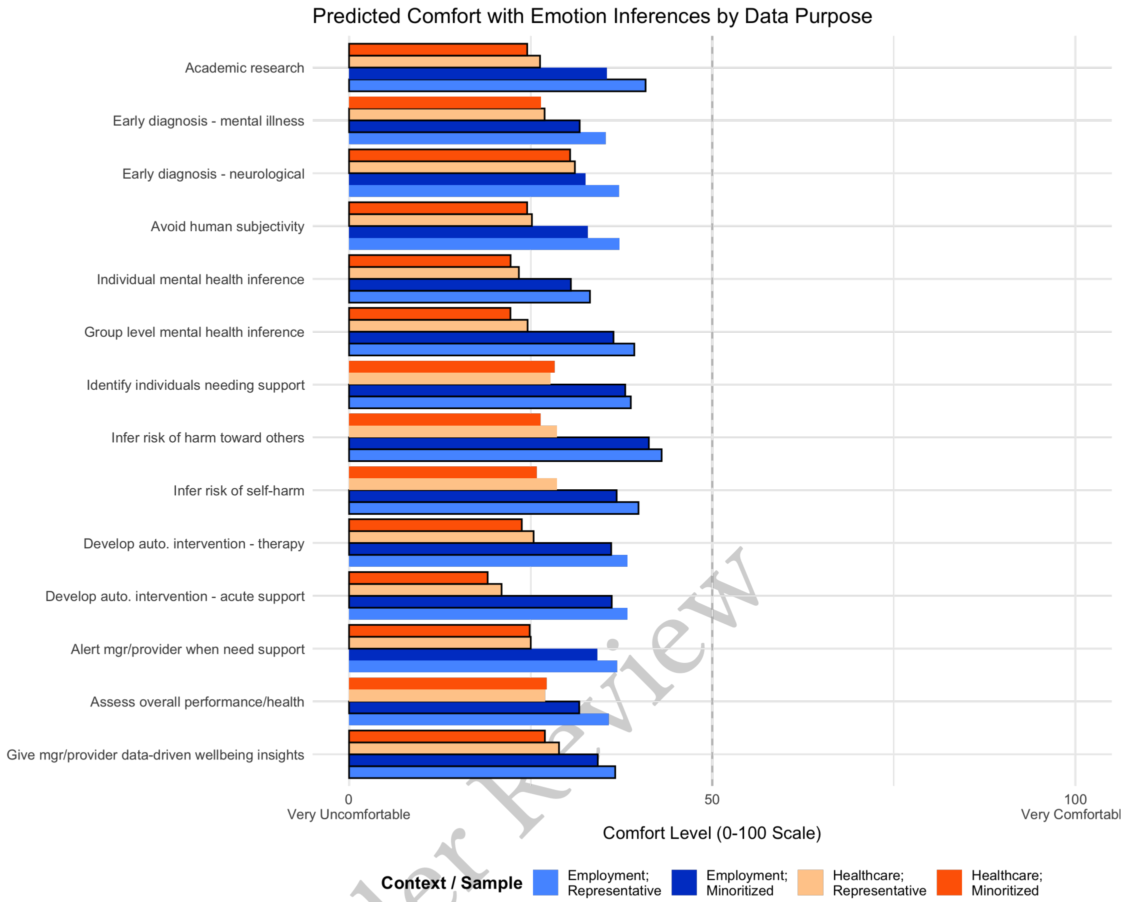

Fig. 3. Predicted Comfort Levels by Purpose. This figure illustrates the predicted comfort levels by combining the purpose variable coefficients to each mixed-effects regression model intercept, derived by analyzing respondent comfort on a scale from 1 (very uncomfortable) to 100 (very comfortable). Bars with black borders indicate statistically significant results.

findings suggest this may reflect greater privacy concerns about employer access to mental health information among marginalized participants [41].

Healthcare context. In healthcare, the negative effect was more pronounced and statistically significant only in the representative sample (representative: $\beta = -1.98$, $SE = 0.98$, $p < 0.05$; minoritized: $\beta = -0.55$, $SE = 0.95$, not significant). The smaller effect in the minoritized group may relate to disparities in mental healthcare quality. Participants from minoritized backgrounds reported difficulties getting providers to recognize their mental health concerns and noted that emotion inferences might help legitimate issues that might otherwise be ignored [143]. Nonetheless, predicted comfort remained low in both samples (26.41 for the minoritized sample and 26.93 for the representative sample), and the difference between them was marginal.

Notably, although early mental health diagnosis aligns with healthcare's broader goals, participants still viewed this purpose as diminishing their emotional privacy.

Earlier diagnosis of neurological disorders. By contrast, using emotion inferences to detect *neurological* disorders had markedly different effects.

Healthcare context. In healthcare, this purpose produced a significantly positive effect on comfort in both samples (representative: $\beta = 2.19$, $SE = 0.98$, $p < 0.05$; minoritized: $\beta = 3.47$, $SE = 0.95$, $p < 0.001$). It was the *only* purpose across all fourteen tested to have a significant positive effect on patient comfort. Despite generally low comfort with emotion inferences, participants viewed this use case as a limited exception that had a positive effect on emotional privacy judgments.

Although our qualitative data did not explicitly address this finding, it may reflect the greater availability of objective measures in neurological diagnostics (e.g., imaging, neurological exams), which could reduce perceived risks to patient autonomy compared to subjective mental health assessments. The stronger positive effect in the minoritized sample suggests these participants may perceive greater potential benefits—or lower risks—from this specific application. However, estimated comfort remained low overall (30.43 for the minoritized sample and 31.1 for the representative sample).

Employment context. In employment, effects were mixed. This purpose had no significant effect in the representative sample ($\beta = 0.55$, $SE = 0.93$, not significant) but showed a weakly significant negative effect in the minoritized sample ($\beta = -1.70$, $SE = 0.89$, $p < 0.1$). The larger and significant negative effect in the minoritized sample suggests workers from marginalized backgrounds perceived higher risks—or fewer benefits—from employer use of emotion inferences for neurological diagnosis.

Coefficient plots (Table **??**) show that, within a 95% interval, the effect for the representative sample crossed zero, while the effect for the minoritized sample did not. This indicates that although the direction of the effect is uncertain for representative participants, it can be confidently interpreted as negative for minoritized workers.

Qualitative data from minoritized participants underscore this discomfort, citing fears of negative personal and professional consequences tied to health disclosures in the workplace [41].

Employee and Patient Assessments. Scholars and technologists have proposed automatic emotion inferences as a potentially objective method to reduce bias in both employee [144] and patient [190] assessments. Rather than relying on self-reports or human observations, incorporating presumably objective emotion inferences into work performance evaluations and health assessments is thought to minimize human subjectivity and the biases involved in understanding individuals' emotional states and their relation to overall work performance and health. We examined how purposes related to augmenting employee and patient assessments influenced participants' comfort with emotion inferences. Participants rated their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable") with their employers or healthcare providers inferring their emotions for the following purposes:

> *(5) avoiding subjectivity in other methods of your employer/healthcare provider learning about your emotional state, like a survey or your employer/healthcare provider's observations*; and *(14) assessing the work performance/overall health of individual employees/patients.*

Predicted comfort levels for these purposes were generally low. On a scale from 0–100, worker comfort ranged from 31.69 to 37.21, while patient comfort ranged from 24.52 to 27.19 (see Figure 3).

Employment context. Employers using emotion inferences to assess overall work performance had a negative effect on worker comfort compared to the baseline purpose, with significance observed only in the minoritized sample (representative: $\beta = -0.88$, $SE = 0.93$, insignificant; minoritized: $\beta = -2.55$, $SE = 0.89$, $p < 0.001$). The larger and significant negative effect in the minoritized sample likely reflects both the general trend in our results—that this sample perceives greater invasions to emotional privacy—and, possibly, increased statistical power from the larger sample size. Qualitative insights from our adjacent studies suggest this discomfort may also reflect concerns among minoritized participants that emotional surveillance could impair work performance or lead to negative employment outcomes [41].

Employers using emotion inferences to reduce subjectivity in understanding workers' emotional states did not yield statistically significant effects in either sample. Taken together, these findings suggest that workers view employer use of emotion inferences for performance assessments as negatively affecting emotional privacy.

Healthcare context. Healthcare providers using emotion inferences to avoid human subjectivity in evaluating patients' emotional states had a significant negative effect on patient comfort in both samples (representative: $\beta = -3.73$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -2.44$, $SE = 0.95$, $p < 0.01$). Although patient comfort was lower in the minoritized sample (24.52 vs. 25.18), the smaller and less significant negative effect in this sample suggests that people of color, individuals with mental illness, and/or minoritized genders may associate relatively higher benefit or reduced risk with this purpose. Our qualitative findings support this interpretation: minoritized participants expressed a desire for more objective, less biased evaluations of their emotional and mental health but also voiced concerns that algorithmic inferences could exacerbate provider bias in practice [143].

By contrast, healthcare providers using emotion inferences to assess overall patient health had no statistically significant effect in either sample, though a weakly significant negative effect (at the $p < 0.1$ level) was observed in the representative sample.

In summary, these results indicate that workers judged employer use of emotion inferences for performance assessments as negatively affecting their emotional privacy. Similarly, patients judged healthcare providers' use of emotion inferences to avoid subjectivity in emotional evaluations as negatively affecting their emotional privacy.

Inferring Mental Health at Individual and Group Levels. We examined how participant comfort was affected by employers and healthcare providers using emotion inferences for the purpose of inferring workers' and patients' mental health at both individual and group levels. Participants were asked to rate their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable") with their employers/healthcare providers inferring their emotions using various data inputs for the purposes of:

(6) inferring the mental health state of employees/patients individually; and

(7) inferring the mental health state of employees/patients. Inferences of an individual's mental health will not be made; only at a group level.

As Figure 3 illustrates, predicted comfort levels for both purposes were low across contexts and samples, with worker comfort ranging from 30.54—39.27 and patient comfort ranging from 22.22—24.59. In both contexts, predicted comfort levels were lower in the marginalized sample than in the U.S. representative sample.

Employment context. Employers using emotion inferences to infer individual workers' mental health had a significant negative effect on comfort in both samples, compared to the baseline purpose (representative: $\beta = -3.48$, $SE = 0.93$, $p < 0.001$; minoritized: $\beta = -3.70$, $SE = 0.89$, $p < 0.001$), with a slightly more pronounced negative effect in the marginalized sample. By contrast, employers inferring workers' mental health at a group level had a significantly positive effect on comfort in both samples (representative: $\beta = 2.63$, $SE = 0.93$, $p < 0.01$; minoritized: $\beta = 2.16$, $SE = 0.89$, $p < 0.05$), with a somewhat smaller effect in the marginalized sample. These results indicate that individual-level mental health inferences are discomforting and perceived as privacy invasive, whereas group-level inferences may be welcomed and seen as relatively privacy-preserving. Consistent with these results, participants in our adjacent qualitative study expressed enthusiasm for the potential of emotion AI to improve mental health support by helping employers identify workplace improvements or resources, tempered by deep concerns about misuse of individual-level emotional information—especially the risk of negative employment outcomes such as termination or loss of opportunities [41]. Our regression results suggest that aggregating emotion inferences may mitigate risks linked to identifiability, balancing perceived benefits with protections for workers' emotional privacy.

Healthcare context. In contrast, healthcare providers using emotion inferences to infer patients' mental health—at either the individual level (representative: $\beta = -5.52$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -4.72$, $SE = 0.95$, $p < 0.001$) or the group level (representative: $\beta = -4.32$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -4.74$, $SE = 0.95$, $p < 0.001$)—had a significant negative impact on patient comfort across both samples. The negative effect of individual-level inferences was slightly smaller in the minoritized sample than in the U.S. representative sample; the effects for group-level inferences were similar across samples. These results indicate that patients are significantly discomforted by mental health inferences regardless of identifiability. Qualitative insights from our related study provide explanatory context: patients expressed concerns that emotion AI could facilitate or worsen harmful mental healthcare practices, such as biased assessments, reduced patient voice, strained provider interactions, and misuse of sensitive information at both the individual and collective levels [143]. Notably, the positive effect associated with group-level inferences observed in the employment context was absent in healthcare. Although our qualitative data did not directly explain this pattern, we suggest that the inherently individualized nature of the patient-provider relationship may account for participants' reluctance to view group-level inferences as alleviating privacy concerns in healthcare.

Societal and Collective Benefit. We examined how employers and healthcare providers using emotion inferences for purposes of societal or collective benefit—specifically, to benefit society by supporting academic research and to benefit workers and patients by identifying individuals requiring additional support to improve mental healthcare resource planning—affected participants' comfort with emotion inferences. Participants rated their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable") with their employers/healthcare providers inferring their emotions using various data inputs for the following purposes:

> *(2) sharing that information with academic researchers to help them learn more about mental health, as part of a research partnership*; and
>
> *(8) identifying employees/patients in need of mental health support, to better plan organizational mental health resources*

Predicted comfort levels remained low overall. As shown in Figure 3, patient comfort (ranging from 24.53—28.3) was consistently lower than worker comfort (ranging from 35.5—40.82) across both purposes and samples.

Employment context. Employers using emotion inferences to support academic research had a positive effect on worker comfort relative to the baseline purpose, with larger and statistically significant effects in the U.S. representative sample only (representative: $\beta = 4.18$, $SE = 0.93$, $p < 0.001$; minoritized: $\beta = 1.26$, $SE = 0.89$, insignificant). While our adjacent study [41] did not surface specific insights explaining this pattern, the result aligns with prior qualitative work indicating that while people hold predominantly negative views toward automatic emotion recognition, their attitudes are less negative in specific use cases involving societal benefit, such as supporting academic research [8, 142]. The positive effect was significantly larger in the representative sample than in the minoritized sample, with a Z-score of 2.38, possibly reflecting heightened mistrust of academic research in minoritized communities due to historical patterns of exclusion and mistreatment [25].

For the purpose of identifying individuals in need of mental health support to inform organizational planning, this use case had a significantly positive impact on worker comfort relative to the baseline in both samples, with a larger effect in the minoritized sample (representative: $\beta = 2.15$, $SE = 0.93$, $p < 0.05$; minoritized: $\beta = 3.78$, $SE = 0.89$, $p < 0.001$). In contrast to the negative effects observed for individual-level emotion inferences in Section 4.1.3, this result suggests that workers' discomfort can be mitigated when emotion inferences are used for purposes that do not assess individual mental health states directly and are instead linked to collective worker benefit. Qualitative results from our related study support this interpretation: nearly one-third of participants, most with marginalized identities, acknowledged potential benefits of using emotion inferences to improve organizational mental health resources and accommodations [41].

Overall, these results suggest that inferences of worker emotion—when used strictly for societal or collective worker benefit—may represent a limited acceptable use case that positively influences emotional privacy judgments. However, sample-level differences also underscore the importance of nuanced, personalized approaches to collecting, using, and sharing emotion inferences that respect diverse privacy needs and preferences.

Healthcare context. By contrast, purposes framed as societal or collective benefit did not preserve patient emotional privacy. Healthcare providers sharing emotion inferences with academic researchers had a significantly negative effect on patient comfort in both samples (representative: $\beta = -2.61$, $SE = 0.98$, $p < 0.01$; minoritized: $\beta = -2.43$, $SE = 0.95$, $p < 0.05$). For the purpose of identifying patients needing support to inform mental healthcare resource planning, the results were not statistically significant at the .05 threshold; however, sample comparisons revealed a statistically significant difference, with a negative effect in the representative sample and a comparatively positive effect in the minoritized sample (Z = -1.84).

Qualitative insights from our related study help explain these patterns. Participants, including many with marginalized identities, acknowledged the potential value of emotion inferences for advancing mental health research. However, they expressed strong concerns about data sharing practices, particularly fears that sharing inferred emotional information with third parties could compromise privacy and create barriers to care [143]. Additionally, higher expectations of confidentiality in the patient-provider relationship likely contributed to the negative effect of this purpose, in contrast to the more positive evaluations observed in employment.

Taken together, these results indicate that even when framed as benefiting society or patients collectively, sharing patient emotion inferences is discomforting and perceived as violating emotional privacy.

Harm Prevention. We investigated whether and how employers and healthcare providers inferring workers' and patients' emotions for the purpose of preventing self-harm and harm toward others influenced comfort with emotion inferences. Participants rated their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable") with their employers/healthcare providers inferring their emotions using various data inputs for the following purposes:

(6) inferring whether employees/patients are at risk of harming themselves; and
(7) inferring whether employees/patients are at risk of harming others

Predicted comfort levels for harm prevention were consistently lower in the healthcare context (25.86—28.65) than in the employment context (36.84—43.03), with lower estimates observed in the minoritized sample than in the U.S. representative sample, as shown in Figure 3.

Employment context. Employers' use of emotion inferences for harm prevention had a significantly positive effect on worker comfort relative to the baseline purpose. Notably, inferring risk of harm toward others had the largest positive effect on worker comfort of any purpose tested in both samples (representative: $\beta = 6.39$, $SE = 0.93$, $p < 0.001$; minoritized: $\beta = 7.03$, $SE = 0.89$, $p < 0.001$). Employers using emotion inferences to infer self-harm also had a significantly positive effect in both samples (representative: $\beta = 3.20$, $SE = 0.93$, $p < 0.01$; minoritized: $\beta = 2.60$,

$SE = 0.89$, $p < 0.01$). Positive effects were similar between samples, suggesting that workers may view employer use of emotion inferences as acceptable for harm prevention purposes, provided that use is limited and justified.

Our adjacent qualitative study offers a nuanced interpretation of these findings. While some workers expressed support for monitoring employee emotions to prevent workplace violence or self-harm—acknowledging potential safety benefits—they emphasized that this would only be acceptable if emotion inferences were restricted strictly to this purpose and proven accurate. Participants expressed deep concern that employers might repurpose emotion inferences for unrelated or punitive uses, or that biased or inaccurate inferences could lead to false flags and unwarranted interventions, ultimately compromising worker safety rather than protecting it [41]. These findings underscore the importance of weighing the potential safety benefits of harm prevention against the serious risks of misuse and error.

Healthcare context. In contrast to the positive effects observed in employment, we found no statistically significant effects of harm prevention purposes on patient comfort in either sample. Trends indicated negative effects for both self-harm (representative: $\beta = -0.27$, $SE = 0.98$, insignificant; minoritized: $\beta = -1.10$, $SE = 0.95$, insignificant) and harm toward others (representative: $\beta = -0.26$, $SE = 0.98$, insignificant; minoritized: $\beta = -0.56$, $SE = 0.95$, insignificant).

Our qualitative analysis provides explanatory insight. Most participants did not perceive benefits to healthcare providers using emotion inferences for harm prevention and expressed substantial privacy concerns. They feared that such uses could legitimize over-surveillance of already vulnerable mentally ill patients and worried that inaccurate inferences could lead to severe consequences, such as unwarranted coercive interventions or involuntary commitment [143].

Supportive Interventions. We examined how emotion inferences used for supportive interventions influenced participants' comfort with emotion inferences in workplace and healthcare contexts. Participants rated their comfort (from 0 = "very uncomfortable" to 100 = "very comfortable") with employers and healthcare providers inferring their emotions using various data inputs for the following purposes:

> *(11) developing an intelligent computer program, such as a chatbot, that can conduct mental health therapy with employees/patients, including you*; *(12) inferring moments employees/patients may be in need of emotional support and responding with an intelligent computer program designed to help employees/patients improve their well-being, such as offering well-being tips*; and *(13) automatically alerting your employer/healthcare provider when employees/patients may need support, including you.*

Predicted comfort levels for these purposes were generally low. Worker comfort ranged from 34.19—38.32 and patient comfort ranged from 19.08—25.42, with lower levels observed in the healthcare context and the minoritized sample (Figure 3).

Employment context. Emotion inferences for supportive interventions had a positive impact on worker comfort when used to develop (representative: $\beta = 1.68$, $SE = 0.93$, $p < 0.1$; minoritized: $\beta = 1.92$, $SE = 0.89$, $p < 0.05$) and deliver (representative: $\beta = 1.66$, $SE = 0.93$, $p < 0.1$; minoritized: $\beta = 1.85$, $SE = 0.89$, $p < 0.1$) automated interventions that provided direct support, relative to the baseline purpose. However, these effects were only weakly significant. Effects were similar between samples. Our analysis did not identify a statistically significant effect for interventions involving third-party alerts to managers or employers.

These results suggest that workers may perceive potential benefits from emotion inferences used to develop or deliver direct well-being interventions—especially when such interventions remain private and do not involve employer oversight. Our qualitative study did not yield direct insights into this specific finding. However, workers expressed a general desire for improved well-being support while also voicing concerns that employer access to inferred emotional information could lead to negative personal and professional consequences [41]. This suggests that workers may cautiously welcome automated well-being interventions provided they protect privacy and are not shared with employers or third parties.

Healthcare context. In contrast, healthcare providers using emotion inferences for supportive interventions had a significantly negative and substantially larger impact on patient comfort across all three purposes. Developing automated mental health therapy had the largest negative effect of any purpose tested (representative: $\beta = -7.91$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -7.88$, $SE = 0.95$, $p < 0.001$). Delivering acute well-being support, such as well-being tips, also had a significant negative effect (representative: $\beta = -3.49$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -3.18$, $SE = 0.95$, $p < 0.001$). Automatically alerting a healthcare provider when support was needed had a significant negative effect as well (representative: $\beta = -3.89$, $SE = 0.98$, $p < 0.001$; minoritized: $\beta = -2.09$, $SE = 0.95$, $p < 0.05$).

Our qualitative analysis suggests several factors contributing to this discomfort. Participants expressed concern that automated well-being interventions could harm patients' mental health through inaccurate inferences or inadequate responses, reduce human interaction between patients and providers, lower the quality of mental healthcare, and breach confidentiality—particularly troubling in a healthcare context characterized by strong expectations for privacy [143].

4.2 Socio-demographic Factors Have Nuanced Influence on Emotional Privacy Judgments of Emotion Inferences in the Workplace and Healthcare

We examined the effect of socio-demographic factors on participants' comfort with emotion inferences in employment and healthcare, specifically race/ethnicity, gender, mental health status, and educational attainment as described in Section 3.2.3 and justified in Section 2.2.

4.2.1 Race/Ethnicity. Compared to white participants, Black participants reported higher comfort with emotion inferences in both employment and healthcare contexts. In employment, this effect was statistically significant in the minoritized sample (representative: $\beta = 5.64$, $SE = 3.19$, $p < 0.1$; minoritized: $\beta = 7.38$, $SE = 2.79$, $p < 0.01$). In healthcare, higher comfort was significant in both samples (representative: $\beta = 3.33$, $SE = 5.44$, $p < 0.01$; minoritized: $\beta = 6.66$, $SE = 3.05$, $p < 0.05$).

Asian participants reported lower comfort with employers inferring their emotions compared to white participants, particularly in the minoritized sample where the effect was statistically significant (representative: $\beta = -3.15$, $SE = 4.31$, insignificant; minoritized: $\beta = -8.05$, $SE = 3.55$, $p < 0.05$). We did not observe significant race/ethnicity effects for Latine or other categories, and no significant effects for Asian participants in the healthcare context.

Across all race/ethnicity categories, Black participants reported the highest comfort levels in both contexts, while Asian participants reported the lowest comfort in employment and white participants reported the lowest comfort in healthcare.

While higher comfort among Black participants may appear surprising given the documented racial and cultural biases present in emotion recognition datasets—leading to potential harms through inaccuracy or discriminatory use [75, 139, 187]—this group may attribute greater potential benefits to emotion inferences or perceive lower risk. Our qualitative studies [41, 143] provide some support for this interpretation. Black participants often highlighted the potential for emotion AI to mitigate racial discrimination and improve emotional support in both employment and healthcare. Yet, they also expressed concern about the risk of perpetuating existing inequities. Importantly, our qualitative analyses did not explicitly investigate the influence of race/ethnicity on participants' perceived risks and

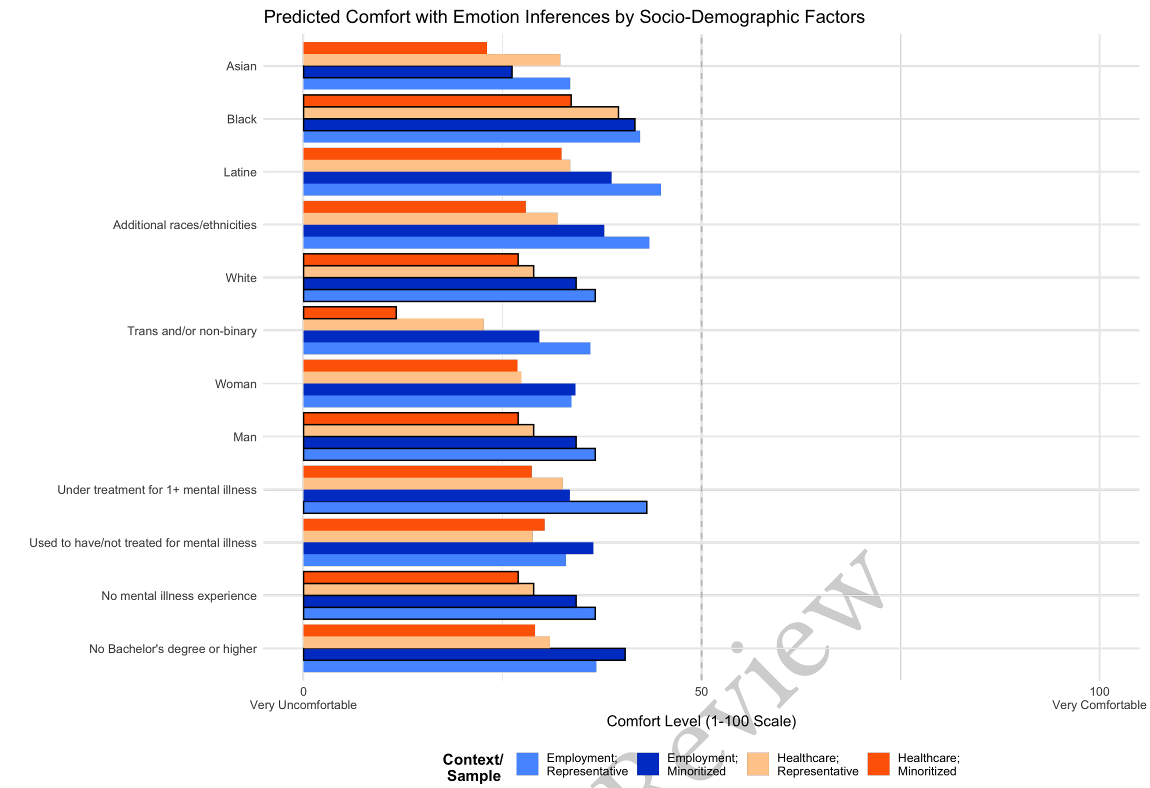

Fig. 4. Predicted Comfort Levels by Socio-demographics. This figure illustrates the predicted comfort levels by combining the socio-demographic variable coefficients to each mixed-effects regression model intercept, derived by analyzing respondent comfort on a scale from 1 (very uncomfortable) to 100 (very comfortable). Bars with black borders indicate statistically significant results.

benefits. Future work is needed to understand how Black workers' and patients' nuanced perspectives on emotion AI shape their emotional privacy judgments.

4.2.2 Gender. In employment scenarios, we did not observe a statistically significant influence for any gender category on participants' comfort in either sample. While prior work suggests that privacy perceptions are often gendered, including in workplace contexts [163], larger sample sizes may be needed to confirm whether gender meaningfully influences emotional privacy judgments concerning emotion inferences in employment and healthcare.

In the healthcare context, however, trans and/or non-binary participants reported significantly less comfort than men on average, with this trend confirmed in the minoritized sample, which included a larger number of trans and/or non-binary participants (representative: $\beta = -6.26$, $SE = 10.99$, insignificant; minoritized: $\beta = -15.32$, $SE = 4.55$, $p < 0.001$). No statistically significant differences were observed for women compared to men in either sample.

Notably, the discomfort reported by trans and/or non-binary participants regarding healthcare providers' use of emotion inferences represents the largest negative effect observed for any socio-demographic factor in our analysis, underscoring substantial emotional privacy concerns about healthcare applications of emotion AI in this group.

4.2.3 Mental Health Status. In the employment context, participants currently under treatment for one or more mental illnesses reported significantly higher comfort with emotion inferences compared to participants with no mental illness, but only in the U.S. representative sample (representative: $\beta = 6.47$, $SE = 3.13$, $p < 0.01$; minoritized: $\beta = -0.79$, $SE = 3.06$, insignificant). While minoritized participants currently under treatment reported lower comfort on average, the result was not statistically significant. As the coefficient range in Table 8 shows, the direction of this variable's impact remains inconclusive in the minoritized sample.

We did not observe statistically significant differences in comfort for participants with resolved or untreated mental illness in either sample.

In the healthcare context, no statistically significant effects were found for any level of mental health status.

The significantly higher comfort observed among participants currently receiving mental health treatment in the U.S. representative sample—but not in the minoritized sample, which included a comparatively higher proportion of such participants—suggests a complex relationship between mental health status and emotional privacy judgments that may vary by intersectional identities. Since our sampling did not differentiate based on specific mental health diagnoses, we recommend future research explore perceptions of emotion inferences among people with particular mental illnesses to better understand and address these perspectives.

4.2.4 Educational Attainment. Compared to participants with a Bachelor's degree or higher, those without a Bachelor's degree reported, on average, higher levels of comfort with emotion inferences across both contexts and samples.

For employer use of emotion inferences, participants without a Bachelor's degree reported higher comfort in both samples. This relationship reached statistical significance only in the minoritized sample, where the positive effect size was substantially larger—a difference likely influenced by the minoritized sample's greater representation of participants without a Bachelor's degree (representative: $\beta = 0.14$, $SE = 2.46$, insignificant; minoritized: $\beta = 6.16$, $SE = 2.37$, $p < 0.01$). In the healthcare context, the relationship between lower educational attainment and comfort with emotion inferences was positive but statistically insignificant in both samples.

The consistently higher comfort reported by participants with lower educational attainment, especially in the employment context, suggests that this group may be less likely to recognize potential risks associated with emotion inferences and/or may perceive greater potential benefits. More research is needed to better understand how educational attainment shapes emotional privacy judgments and risk-benefit perceptions related to emotion AI.

4.3 Individual Privacy Belief Factors Affect Emotional Privacy Judgments When Contextualized

We investigated whether and how individual privacy beliefs—including general privacy concerns, trust in employers' and healthcare providers' handling of sensitive information, and perceived sensitivity of emotional information—affected participants' comfort with emotion inferences.

4.3.1 General Privacy Concerns. Participants' level of general privacy concerns did not have a statistically significant effect on their comfort with emotion inferences in either context or sample.

4.3.2 Trust in Employers' and Healthcare Providers' Handling of Sensitive Information. The level of trust participants attributed to their employers' and healthcare providers' handling of their sensitive information significantly and positively influenced their comfort with emotion inferences in both contexts. Participants reporting higher levels of trust reported significantly higher comfort with emotion inferences in both employment (representative: $\beta = 0.54$,

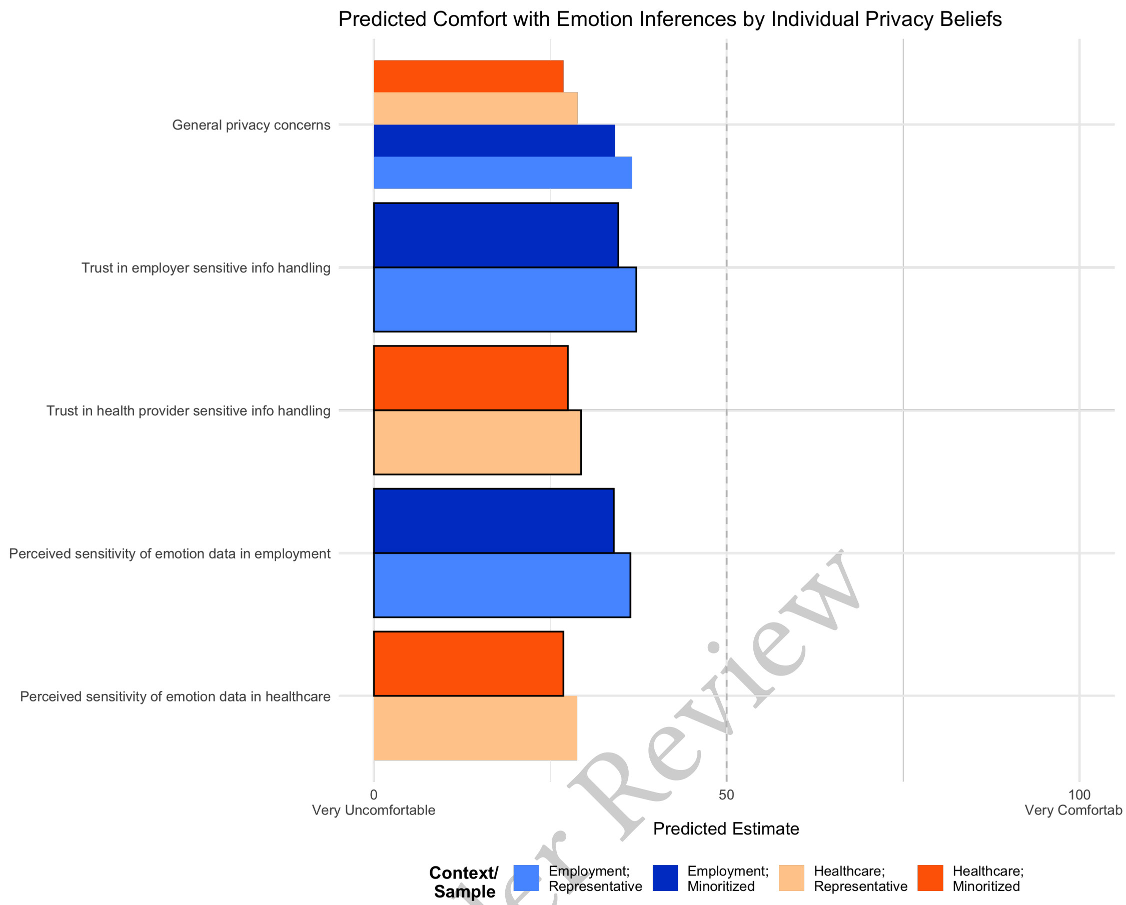

Fig. 5. Predicted Comfort Levels by Individual Privacy Beliefs. This figure illustrates the predicted comfort levels by combining the individual privacy belief variable coefficients with each mixed-effects regression model intercept, derived by analyzing respondent comfort on a scale from 1 (very uncomfortable) to 100 (very comfortable). Bars with black borders indicate statistically significant results.

$SE = 0.05$, $p < 0.001$; minoritized: ($\beta = 0.40$, $SE = 0.05$, $p < 0.001$) and healthcare (representative: $\beta = 0.44$, $SE = 0.08$, $p < 0.001$; minoritized: ($\beta = 0.53$, $SE = 0.05$, $p < 0.001$) contexts.

Of note, this effect was significantly different between samples for the healthcare context; the Z-score of 2.09 indicates that positive trust beliefs had a greater influence on patient comfort in the U.S. representative sample than in the minoritized sample.

4.3.3 Perceived Sensitivity of Emotional Information. Participants rated the level of sensitivity they associated with emotional information along with other information types already categorized in law and literature as sensitive – political opinions, religious beliefs, biometric data, health information, sex life/sexual orientation, genetic information, and union membership [38, 134] – when handled by one's employer and healthcare provider. As participants answered this question in a post-test after responding to vignettes that described various uses of their emotion inferences, we expect that responses are indicative of participants' perceptions of emotion inferences.

Employment context. As the box plot in Figure 6 illustrates, participants rated the sensitivity of emotional information handled by one's employer similar to data types already recognized as sensitive. The median level of perceived sensitivity of emotional information handled by employers for participants in the representative sample ranks higher than that for genetic information, health information, and union membership. The median sensitivity rating for emotional information handled by employers in the minoritized sample ranked among the lowest of sensitive data types, with a similar sensitivity to political opinions.

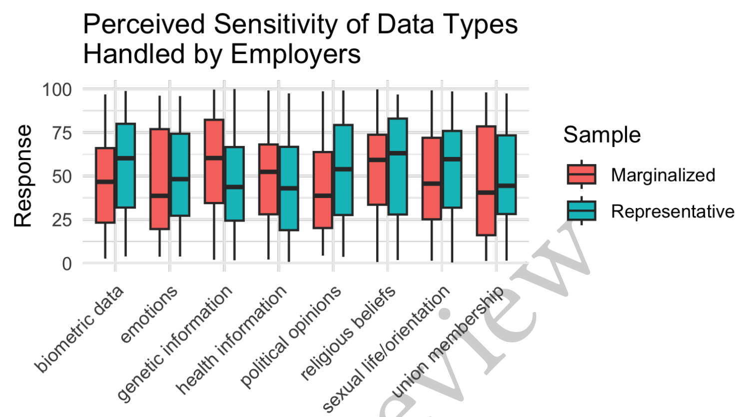

Fig. 6. Perceived sensitivity ratings of emotional information compared to other sensitive data types in the employment context. Box plots show the distribution of sensitivity ratings for each data type by sample, on a scale from 1 (not sensitive) to 100 (extremely sensitive).

Healthcare context. Participants rated the sensitivity of emotional information handled by healthcare providers higher than when handled by employers, as shown in Figure 7. Participants in the representative sample perceived the sensitivity of emotional information handled by healthcare providers higher than biometric data, health information, political opinions, religious beliefs, and union membership. In contrast to their relatively lower perceived sensitivity of emotion data information handled by employers compared to other data types, participants in the minoritized sample rated emotion data information's sensitivity higher when handled by healthcare providers than all other sensitive information types.

In addition, our analysis examined whether and how participants' perceived sensitivity of emotional information when handled by employers and healthcare providers affected their comfort with emotion inferences. We found that participants' perceived sensitivity of emotion data had a significant effect on their comfort with emotion inferences in both contexts. Participants associating emotional information with higher sensitivity reported significantly less comfort with emotion inferences in the employment context (representative: $\beta = -0.30$, $SE = 0.05$, $p < 0.001$; minoritized: ($\beta = -0.25$, $SE = 0.05$, $p < 0.001$). Participants similarly reported less comfort with emotion inferences in healthcare (representative: $\beta = -0.10$, $SE = 0.05$, $p < 0.1$; minoritized: ($\beta = -0.11$, $SE = 0.04$, $p < 0.01$) contexts, with a significant effect confirmed in the minoritized sample.

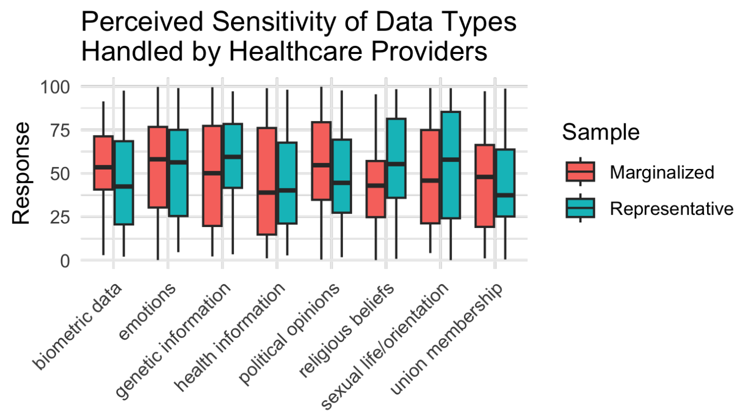

Fig. 7. Perceived sensitivity ratings of emotional information compared to other sensitive data types in the healthcare context. Box plots show the distribution of sensitivity ratings for each data type by sample, on a scale from 1 (not sensitive) to 100 (extremely sensitive).

In light of these findings, we bring attention to the fact that personal emotional information is not currently recognized in existing literature or legal/regulatory frameworks as a special category of data that requires heightened protections and safeguards—an issue we elaborate on in the discussion.

5 DISCUSSION

Workers' and patients' ability to protect their *emotional privacy*—managing whether, how, and to what extent information about their emotions is collected, used, and shared—is increasingly threatened by technologies that automatically infer human emotions. Although proponents highlight potential benefits, such as earlier medical diagnoses, harm prevention, and enhanced emotional well-being, these claims remain largely speculative and scientifically unvalidated. Meanwhile, the automatic inference of emotional information poses significant privacy risks, particularly in the U.S. as powerful actors like employers and healthcare providers adopt these technologies with minimal oversight.

As emotion AI development and regulation continues to evolve, the privacy perceptions of those subject to emotion inference should help shape ethical standards. Emotional privacy is not simply about disclosure; it reflects contextually-dependent norms [124] governing how emotional information is collected and used. It is a matter of degree [140]: while emotions may be perceptible through facial expressions, language, or vocal patterns, this does not justify unrestricted automatic inference. Emotion inference technologies expand the set of contextual actors—both human and non-human—who can access, use, and share individuals' emotional information in opaque ways. Individuals should have a meaningful capacity to decide whether, how, and to what extent their emotional information is inferred and used.

Protecting emotional privacy is essential not only for respecting human dignity and promoting individual well-being. Privacy and agency in emotion-inference technologies are critical to sustaining institutional trust and enabling people to engage with such systems without undue risks. For scholars concerned with justice and human values in socio-technical systems, it is vital to understand how to protect emotional privacy in ways that reflect the factors shaping privacy judgments—particularly among those most vulnerable to harm.

Our findings contribute to this goal by empirically identifying contextual and identity-based vulnerabilities that influence emotional privacy judgments. These vulnerabilities are especially pronounced among minoritized groups, whose emotional privacy needs and concerns can differ markedly from dominant groups. Relying solely on socially dominant "internal standards of justice" [131] to define privacy norms risks reinforcing systemic injustices and silencing dissenting views [26, 114, 124, 181]. Aligning the development, design, and regulation of emotion-inference technologies with the unmet needs and persistent concerns of those most vulnerable to technological impact benefits everyone.

Our results also have significant policy relevance. Notably, the patterns in our study mirror many elements of the EU AI Act and its clarified application guidelines [53]. We observe striking alignment between public intuitions and the regulatory architecture: where the Act imposes strict limits—regulating biometric inputs, prohibiting individual profiling in employment, addressing power asymmetries—participants' comfort drops sharply. Where the Act permits narrow exceptions—workplace safety, neurological monitoring for medical applications, or non-identifiable aggregated insights—comfort increases. This convergence suggests that the distinctions respondents draw closely match the EU's regulatory reasoning, offering regulatory legitimacy to the privacy judgments surfaced in our study.

Beyond reinforcing current regulatory directions, our findings offer a model for anticipating future governance challenges. As commercial innovation adapts to regulatory constraints and as jurisdictions develop more detailed rules, this study provides both empirical evidence to inform those efforts and a methodological approach capable of identifying socially salient privacy boundaries as they evolve. In the sections that follow, we discuss implications for practice, policy, and research, emphasizing:

(1) **inference minimization**—limiting the purposes for collecting or using emotional information
(2) Recognizing emotional information as a **sensitive data category**;
(3) Advancing **contextual and demographic sensitivity** in the design, application, and regulation of emotion-inference technologies.

5.1 Limiting Purpose through Inference Minimization Principles

Contextual integrity assumes data purpose is constrained implicitly—encoded in transmission principles and justified by the context's goals—rather than specified as a stand-alone parameter [124]. Our results show that purpose nonetheless drives emotional privacy judgments. In the workplace, performance-scoring inferences—a common managerial practice designed to boost productivity [5, 14, 76, 192]—elicited *lower* comfort. By contrast, employers sharing workers' emotion inferences with academic researchers—an extraneous purpose that does not obviously advance workplace goals—*raised* comfort levels. A similar pattern surfaced in healthcare. Automating interventions or diagnosing mental illness—purposes tightly coupled to clinical objectives—*lowered* comfort, likely because they override patient-initiated disclosure and undermine interpretive agency. Yet neurological disorder screening, another clinical use, *increased* comfort.

These divergences cannot be explained by contextual integrity's five canonical parameters or by contextual goals alone. Instead, the explicit *purpose* of the inference—together with the type of information, actors involved, and the transmission principles governing the flow—decisively shapes emotional privacy judgments. As Nissenbaum has noted, purpose's relevance to contextual integrity has become more salient alongside evolving technologies and data practices [127], and adding a purpose dimension may be a "necessary [policy] antidote" [126]. These findings support extending contextual integrity to treat purpose as a constitutive contextual parameter, enabling more precise governance through purpose limitation and inference minimization rules. Drawing on the quantitative patterns and complementary qualitative results from this study [41, 143], we propose a narrow, empirically grounded palette of permissible purposes for emotion inference. Key safeguards include:

- **Purpose binding:** Mirroring the EU AI Act's risk-based, context-specific approach [53], define each permissible purpose narrowly and exhaustively (e.g., real-time fatigue detection in safety-critical roles). Specify parallel prohibited purposes (e.g., burnout or depression screening). Any secondary use—or any use outside the narrowly scoped carve-out—should be categorically barred.
- **Granular, opt-in consent:** allow individuals to opt-in or withdraw for each distinct use of emotion data, raw or inferred.
- **Ex ante validation:** require evidence of claimed benefits before deployment
- **Minimal retention:** store only what is strictly necessary for the stated purpose
- **Robust controls:** enforce access limits, encryption, and anonymization, overseen by independent auditors.

Embedding these constraints in law, design, and institutional policy would operationalize contextual integrity's normative commitments, protecting emotional privacy while allowing only narrowly justified, socially valuable uses of emotion AI systems.

5.2 Recognizing the Sensitivity of Emotional Data

Our findings confirm that workers and patients perceive emotion inferences as highly sensitive, often rating them as more sensitive than established categories such as biometric or genetic data. Yet emotion data remains unrecognized as a special category of sensitive information in most privacy frameworks [15].

This sensitivity reflects significant, context-specific risks. In workplaces, emotion inferences could enable discrimination on the basis of perceived mental disability—even without direct disclosure. In healthcare, inaccurate inferences may trigger misdiagnosis or stigma. If exported beyond the original context (e.g., sold to data brokers), such inferences could fuel exploitative advertising or other downstream harms. These participant-voiced concerns [41, 143], together with the strong negative coefficient for perceived sensitivity and the consistently low comfort scores, help to explain why participants regard emotional information as an acutely sensitive data type in both settings. As prior work suggests, privacy concern rises when information heightens vulnerability to harm [103, 108, 146].

Our findings support the formal classification of emotional information as a sensitive category of data. Doing so would require data handlers to apply heightened safeguards [52] aligned with the inference minimization principles we propose above. It would also address persistent concerns expressed by participants about the adequacy of self-regulation in power-imbalanced institutional settings. Sensitivity classification would support regulators in identifying privacy risks and compel both industry and academic practitioners to specify how emotion data is collected, used, and protect.

Finally, such classification would extend urgently needed protections to controversial technologies like facial emotion recognition. Our findings show that the use of facial data consistently heightened discomfort—likely reflecting broader public concerns with facial recognition technologies [191]. Current U.S. regulation typically limits protections to biometric identification [83]. Defining emotional data as sensitive should cover both raw inputs and inferred outputs, closing existing regulatory gaps and better aligning policy with the emotional privacy norms surfaced in our study.

5.3 Advancing Contextual Sensitivity and Positional Vulnerability in Emotional Privacy Research and Practice

Our findings reveal that comfort with emotion inferences varies not only by purpose but also by social position. Although not all socio-demographic effects were statistically significant, several patterns across race, gender, mental health status, and education were illuminating. In both employment and healthcare contexts, Black participants consistently reported higher comfort relative to white participants, with mean comfort levels higher for this group than for any other racial/ethnic category. Similarly, participants without a Bachelor's degree tended to view emotion AI data flows more favorably across the board, including a substantial and statistically significant effect in employment within the minoritized sample (+6.16) compared to a near-zero effect in the representative sample (+0.14). These patterns suggest that *position-related vulnerability* may heighten recognition of when data flows align with the legitimate social ends of a context—such as promoting well-being and support—thereby upholding dignity and fair treatment in the workplace [9] and preserving patient autonomy and dignity in healthcare [177]. Ethical governance of emotional privacy must therefore balance harm prevention with recognition of benefits, particularly for those most vulnerable to harm and exclusion.

At the level of intersecting socio-demographic variables, notable patterns emerged:

- **Trans and/or non-binary participants** reported heightened discomfort, especially toward emotion inferences in healthcare.
- **Participants undergoing treatment for mental illness** judged emotion inferences more positively in healthcare (across both samples), but only in the U.S. representative sample did this translate to the workplace.
- **Participants with untreated or resolved mental illness** expressed more negative judgments in the representative sample, but more positive judgments in the minoritized sample.
- **Asian participants** tended to judge emotion inferences more negatively, especially in the workplace.
- **Black participants and those without a Bachelor's degree** reported consistently higher comfort, significantly so in both employment and healthcare.

We also observed key differences at the belief level. General privacy beliefs, as measured by the Internet Users' Information Privacy Concern (IUIPC) scale, did not significantly predict emotional privacy judgments. Instead, context-specific beliefs (e.g., perceived sensitivity of emotional data and trust in employers or healthcare providers) emerged as decisive predictors. This finding challenges the adequacy of general privacy concern frameworks like IUIPC and underscores the need for research approaches that attend to both contextual and position-based variations.

By identifying how contextual, socio-demographic, and belief-based factors intersect to shape emotional privacy judgments, our findings underscore the importance of designing, applying, and regulating emotion-inference technologies with both contextual and individual sensitivity. While not all observed differences achieved statistical significance—unsurprising given power constraints for some intersecting groups—the patterns nonetheless offer theoretically meaningful insights into how lived experience, privacy vulnerability, and position-based trade-offs shape privacy judgments. Privacy research, too, must move beyond aggregate or nationally representative models to reflect the diverse privacy needs, concerns, and expectations of different people and groups.

Locating risk at the data flow level: a human-centered design implication. Across these patterns, power asymmetries—particularly in the employer/employee and provider/patient relationships—emerged as central to shaping comfort with emotion inferences. Consistent with contextual integrity theory, our findings suggest that emotional privacy concerns are less about the technology itself and more about the institutional contexts and data flows in which it operates. Where emotion inference technologies were perceived as potentially beneficial, participants also voiced concern that institutional power dynamics could undermine agency or lead to harm.

Notably, our findings also revealed important *intra-contextual* distinctions: for example, in healthcare, participants judged emotion inferences for neurological disorders more favorably than for mental health monitoring, reflecting how purpose functions as a critical—yet often overlooked—determinant of appropriateness even within the same domain. This reinforces our empirical extension of contextual integrity by elevating purpose as a constitutive parameter shaping privacy judgments.

These insights point to a clear human-centered design implication. One strategy for mitigating risks while preserving benefits is to remove or limit data flows that embed institutional power asymmetries. Deploying emotion inference technologies in self-monitoring or closed contexts—where individuals retain control and interpretive agency over data capture, use, and sharing—may help protect privacy and promote autonomy. Prior research on participatory and agency-supportive data practices, such as semi-automated self-monitoring systems, shows how design can balance automation with self-determination [35].

Of course, such deployments are only appropriate where the data flows themselves adhere to *contextually appropriate parameters*, as defined by contextual integrity: including suitable transmission principles (e.g., limits on sharing and retention), clearly justified purposes, and alignment with the social norms and goals of the context. This is especially critical as emotion data increasingly circulates across sectors. Related work on cross-sectoral data sharing highlights both the potential benefits and challenges of such practices, including the need for transparency, consent specificity, and recognition of cohort-based risks [117]—all elements that contextual integrity explicitly requires. Our findings suggest that while cross-sectoral sharing may be acceptable when it demonstrably serves participants' goals and adheres to trusted contextual parameters (as in some clinical research settings), default sharing of emotion data beyond the original context remains a major source of discomfort and must be carefully governed. Further research is needed to validate what contextual parameters are judged appropriate across diverse groups and use cases, especially in emerging or hybrid contexts where norms are not yet fully established.

Finally, our study itself reflects a human-centered design approach. By systematically analyzing how people's privacy judgments vary by context, purpose, and social position—and by identifying the specific data flows that drive acceptance or rejection—we demonstrate how empirical, participant-centered methods can inform both technology design and governance.

6 CONCLUSION

Emotion AI technologies introduce unprecedented flows of affective data into domains where privacy, dignity, and well-being are at stake. By testing 56 workplace and healthcare scenarios with two demographically differentiated U.S. samples, we show that *contextual, socio-demographic, and privacy belief* factors jointly shape how workers and patients judge the acceptability of those data flows:

(1) **Purpose dominates.** Varying the stated aim of an emotion inference produces the largest shifts in comfort. Purposes that reinforce a context's social mission (e.g., safety in employment, neurological screening in healthcare) raise comfort; purposes that distort those missions (e.g., performance scoring, automated mental health diagnosis) lower it.

(2) **Input modality matters.** Facial analytics consistently reduce comfort relative to speech/text, reflecting persistent skepticism toward vision-based emotion recognition.
(3) **Position-related vulnerability influences judgments.** Minoritized participants follow the same directional trends as the representative cohort, but with amplified effects—positive and negative—consistent with greater perceived susceptibility to both risks and benefits.
(4) **Belief factors are decisive.** Institutional trust raises comfort; perceived sensitivity of emotional information lowers it—often more than recognized sensitive data categories.

Theoretical contribution. Our findings empirically extend contextual integrity by demonstrating that purpose—traditionally treated as implicit—functions as an inter-dependent, constitutive parameter. Elevating purpose clarifies why otherwise similar data flows diverge in perceived appropriateness and provides a tractable lever for governance.

Design and policy implications.

- **Purpose limitation and inference minimization.** Regulation and policy should enumerate narrowly tailored, validated purposes; bar secondary uses; and require necessity proofs before deployment—mirroring risk-based approaches such as the EU AI Act [53].
- **Elevating emotional data protections.** Emotional information, including inferred emotion, warrants protection as a special category of data with heightened safeguards. Given the predictive power of trust in shaping privacy judgments, design and deployment should embed transparency, auditability, and meaningful opt-out rather than rely on institutional goodwill.
- **Individual sensitivity.** Both contextual and individual factors shape emotional privacy judgments. Privacy research, system design, and governance frameworks addressing emotional privacy-intrusive technologies should therefore explicitly attend to varying susceptibilities and diverse needs by upholding the dignity and agency of all data subjects.

Future work. Longitudinal and qualitative research should trace how emotional privacy judgments evolve with repeated exposure to emotion-inference systems, extend these findings to additional high-stakes domains (e.g., education, law enforcement), and engage affected communities in co-designing technologies that reflect their values, needs, and vulnerabilities. Realizing the potential benefits of emotion AI must not require the indefensible trade-off of sacrificing emotional privacy.

As emotion AI proliferates, its impact on human dignity will depend not only on how sharply we define and enforce the purposes for which emotional data may flow, but also on how effectively we embed vulnerability sensitivity, positional equity, and context-aware design. Purpose-aware extensions to contextual integrity, paired with inference-minimization, sensitivity classification, and participatory design, offer a principled and actionable path forward for researchers, designers, and policymakers.

REFERENCES

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.
[2] Alessandro Acquisti and Jens Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE security & privacy* 3, 1 (2005), 26–33.
[3] Accountability Act. 1996. Health insurance portability and accountability act of 1996. *Public law* 104 (1996), 191.
[4] Idris Adjerid, Alessandro Acquisti, and George Loewenstein. 2019. Choice architecture, framing, and cascaded privacy choices. *Management science* 65, 5 (2019), 2267–2290.
[5] Ifeoma Ajunwa, Kate Crawford, and Jason Schultz. 2017. Limitless worker surveillance. *Calif. L. Rev.* 105 (2017), 735.

[6] I. Elaine Allen and Christopher A. Seaman. 2007. Likert Scales and Data Analyses. *Quality Progress* (July 2007).

[7] Mansour Naser Alraja, Murtaza Mohiuddin Junaid Farooque, and Basel Khashab. 2019. The effect of security, privacy, familiarity, and trust on users' attitudes toward the use of the IoT-based healthcare: the mediation role of risk perception. *Ieee Access* 7 (2019), 111341–111354.

[8] Nazanin Andalibi and Justin Buss. 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[9] Elizabeth Anderson. 2023. *Hijacked: How neoliberalism turned the work ethic against workers and how workers can take it back.* Cambridge University Press.

[10] Ira Anjali Anwar, Joyojeet Pal, and Julie Hui. 2021. Watched, but moving: Platformization of beauty work and its gendered mechanisms of control. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–20.

[11] World Medical Association et al. 2004. Declaration of Helsinki: World Medical Association. *Ethical principles for medical research on human beings* 64 (2004), 2013.

[12] Oshrat Ayalon and Eran Toch. 2017. Not even past: Information aging and temporal privacy in online social networks. *Human–Computer Interaction* 32, 2 (2017), 73–102.

[13] Nazish Azam, Tauqir Ahmad, and Nazeef Ul Haq. 2021. Automatic emotion recognition in healthcare data using supervised machine learning. *PeerJ Computer Science* 7 (2021), e751.

[14] Kirstie Ball, Elizabeth M Daniel, and Chris Stride. 2012. Dimensions of employee privacy: an empirical study. *Information Technology & People* 25, 4 (2012), 376–394.

[15] Jennifer S Bard. 2021. Developing Legal Framework for Regulating Emotion AI. *BUJ Sci. & Tech. L.* 27 (2021), 271.

[16] Kim Bartel Sheehan. 1999. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of interactive marketing* 13, 4 (1999), 24–38.

[17] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).

[18] Howard Beales and Jeffrey A Eisenach. 2013. Putting consumers first: A functionality-based approach to online privacy. *Available at SSRN 2211540* (2013).

[19] Annika Bergström. 2015. Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behavior* 53 (2015), 419–426.

[20] Suze Berkhout and Juveria Zaheer. 2021. Digital Self-Monitoring, Bodied Realities: Re-Casting App-Based Technologies in First Episode Psychosis. *Catalyst: Feminism, Theory, Technoscience* 7, 1 (2021).

[21] Jaspreet Bhatia and Travis D Breaux. 2018. Empirical measurement of perceived privacy risk. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 6 (2018), 1–47.

[22] Phillip A. Bishop and Robert L. Herron. 2015. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *International Journal For Exercise Science* 8, 3 (2015), 297–302.

[23] Johnna Blair, Dahlia Mukherjee, Erika FH Saunders, and Saeed Abdullah. 2023. Knowing How Long a Storm Might Last Makes it Easier to Weather: Exploring Needs and Attitudes Toward a Data-driven and Preemptive Intervention System for Bipolar Disorder. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–12.

[24] Paul D Bliese, Mark A Maltarich, and Jonathan L Hendricks. 2018. Back to basics with mixed-effects models: Nine take-away points. *Journal of Business and Psychology* 33, 1 (2018), 1–23.

[25] Billie Bonevski, Madeleine Randell, Chris Paul, Kathy Chapman, Laura Twyman, Jamie Bryant, Irena Brozek, and Clare Hughes. 2014. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC medical research methodology* 14 (2014), 1–29.

[26] Anne Bowser, Katie Shilton, Jenny Preece, and Elizabeth Warrick. 2017. Accounting for privacy in citizen science: Ethical research in a context of openness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2124–2136.

[27] Karen Boyd and Nazanin Andalibi. 2023. Automated emotion recognition in the workplace: How proposed technologies reveal potential futures of work. *Proceedings of the ACM on human–computer interaction* (2023).

[28] Aaron R Brough and Kelly D Martin. 2020. Critical roles of knowledge and motivation in privacy research. *Current opinion in psychology* 31 (2020), 11–15.

[29] EPM Brouwers, MCW Joosen, C Van Zelst, and J Van Weeghel. 2020. To disclose or not to disclose: a multi-stakeholder focus group study on mental health issues in the work environment. *Journal of occupational rehabilitation* 30 (2020), 84–92.

[30] Simone Browne. 2015. *Dark matters: On the surveillance of blackness.* Duke University Press.

[31] Nicholas P Burnett, Alyssa M Hernandez, Emily E King, Richelle L Tanner, and Kathryn Wilsterman. 2022. A push for inclusive data collection in STEM organizations. *Science* 376, 6588 (2022), 37–39.

[32] Matt Cain and Michael Woodbridge. 2020. Hype Cycle for the Digital Workplace, 2020. https://www.gartner.com/document/3987663?ref=solrAll&refval=286087450

[33] Adam Mourad Chekroud, Ryan Joseph Zotti, Zarrar Shehzad, Ralitza Gueorguieva, Marcia K Johnson, Madhukar H Trivedi, Tyrone D Cannon, John Harrison Krystal, and Philip Robert Corlett. 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* 3, 3 (March 2016), 243–250. https://doi.org/10.1016/S2215-0366(15)00471-X

[34] Carl J Chimi and David L Russell. 2009. The Likert scale: A proposal for improvement using quasi-continuous variables. In *Information Systems Education Conference, Washington, DC*. 1–10.

[35] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A Epstein, Felicia Cordeiro, Matthew Kay, Gregory D Abowd, Tanzeem Choudhury, James Fogarty, et al. 2017. Semi-automated tracking: a balanced approach for self-monitoring applications. *IEEE Pervasive Computing* 16, 1 (2017), 74–84.

[36] Seung Youn Yonnie Chyung, Ieva Swanson, Katherine Roberts, and Andrea Hankinson. 2018. Evidence-Based Survey Design: The Use of Continuous Rating Scales in Surveys. *Performance Improvement* 57, 5 (May 2018), 38–48. https://doi.org/10.1002/pfi.21763

[37] Danielle Keats Citron and Daniel J Solove. 2022. Privacy harms. *BUL Rev.* 102 (2022), 793.

[38] Damian Clifford, Megan Richardson, and Normann Witzleb. 2020. Artificial Intelligence and Sensitive Inferences: New Challenges for Data Protection Laws. *Regulatory Insights on Artificial Intelligence: Research for Policy (Edward Elgar, 2021)* (2020).

[39] Roy Coleman. 2004. Reclaiming the streets: Closed circuit television, neoliberalism and the mystification of social divisions in Liverpool, UK. *Surveillance & Society* 2, 2/3 (2004).

[40] Patricia Hill Collins. 2019. *Intersectionality as critical social theory*. Duke University Press.

[41] Shanley Corvite*, Kat Roemmich*, Tillie Rosenberg, and Nazanin Andalibi. 2023. Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens. *Proceedings of the ACM on human–computer interaction* (2023).

[42] Kaitlin L Costello and Diana Floegel. 2020. "Predictive ads are not doctors": Mental health tracking and technology companies. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e250.

[43] Kimberlé Crenshaw. 2013. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In *Feminist legal theories*. Routledge, 23–51.

[44] Mandar Deshpande and Vignesh Rao. 2017. Depression detection using emotion artificial intelligence. In *2017 international conference on intelligent sustainable systems (iciss)*. IEEE, 858–862.

[45] Marwan Dhuheir, Abdullatif Albaseer, Emna Baccour, Aiman Erbad, Mohamed Abdallah, and Mounir Hamdi. 2021. Emotion recognition for healthcare surveillance systems using neural networks: A survey. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 681–687.

[46] Lisa M Diamond and Jenna Alley. 2022. Rethinking minority stress: A social safety perspective on the health effects of stigma in sexually-diverse and gender-diverse populations. *Neuroscience & Biobehavioral Reviews* 138 (2022), 104720.

[47] Catherine D'ignazio and Lauren F Klein. 2020. *Data feminism*. MIT press.

[48] Sidney K D'mello and Jacqueline Kory. 2015. A review and meta-analysis of multimodal affect detection systems. *ACM computing surveys (CSUR)* 47, 3 (2015), 1–36.

[49] Iris Eekhout, Michiel R de Boer, Jos WR Twisk, Henrica CW de Vet, and Martijn W Heymans. 2012. Brief report: missing data: a systematic review of how they are reported and handled. *Epidemiology* (2012), 729–732.

[50] Marta Elliott and Jordan C Reuter. 2021. Disclosure, discrimination, and identity among working professionals with bipolar disorder or major depression. (2021).

[51] Darren Ellis and Ian Tucker. 2020. *Emotion in the digital age: Technologies, data and psychosocial life*. Routledge.

[52] Amitai Etzioni. 2014. A cyber age privacy doctrine: More coherent, less subjective, and operational. *Brook. L. Rev.* 80 (2014), 1263.

[53] European Commission. 2025. *Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689 (AI Act)*. Communication to the Commission C(2025) 884 final. European Commission, Brussels. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act Draft guidelines annexed to the Communication; formally adopted upon.

[54] W Holmes Finch, Jocelyn E Bolin, and Ken Kelley. 2019. *Multilevel modeling using R*. Crc Press.

[55] Melissa L Finucane, Ali Alhakami, Paul Slovic, and Stephen M Johnson. 2000. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making* 13, 1 (2000), 1–17.

[56] Annette Flanagin, Tracy Frey, Stacy L Christiansen, and Howard Bauchner. 2021. The reporting of race and ethnicity in medical and science journals: comments invited. *Jama* 325, 11 (2021), 1049–1052.

[57] Joshua Fogel and Elham Nehmad. 2009. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior* 25, 1 (2009), 153–160.

[58] United States. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1978. *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. Vol. 2. Department of Health, Education, and Welfare, National Commission for the ....

[59] Nikolaus Forgó, Stefanie Hänold, and Benjamin Schütze. 2017. The principle of purpose limitation and big data. *New technology, big data and the law* (2017), 17–42.

[60] Michel Foucault. 2012. *Discipline and punish: The birth of the prison*. Vintage.

[61] Max Freyd. 1923. The graphic rating scale. *Journal of educational psychology* 14, 2 (1923), 83.

[62] Deepak Garg, Limin Jia, and Anupam Datta. 2011. Policy auditing over incomplete logs: theory, implementation and applications. In *Proceedings of the 18th ACM conference on Computer and communications security*. 151–162.

[63] Vaibhav Garg, Lesa Lorenzen-Huber, L Jean Camp, and Kay Connelly. 2012. Risk communication design for older adults. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 29. IAARC Publications, 1.

[64] Andrew Gelman and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models.* Cambridge university press.

[65] Luana Giatti, Lidyane do Valle Camelo, Jôsi Fernandes de Castro Rodrigues, and Sandhi Maria Barreto. 2012. Reliability of the MacArthur scale of subjective social status-Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *BMC public health* 12, 1 (2012), 1–7.

[66] Gregorio González-Alcaide, Mercedes Fernández-Ríos, Rosa Redolat, and Emilia Serra. 2021. Research on Emotion Recognition and Dementias: Foundations and Prospects. *Journal of Alzheimer's Disease* 82, 3 (Aug. 2021), 939–950. https://doi.org/10.3233/JAD-210096

[67] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H Mhaidli, Kentaro Toyama, and Florian Schaub. 2018. Keeping a low profile? Technology, risk and privacy among undocumented immigrants. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–15.

[68] Oliver L Haimson. 2019. Mapping gender transition sentiment patterns via social media data: toward decreasing transgender mental health disparities. *Journal of the American Medical Informatics Association* 26, 8-9 (2019), 749–758.

[69] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 466:1–466:35. https://doi.org/10.1145/3479610

[70] Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M Branham. 2018. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–13.

[71] Paul Harpur, Fitore Hyseni, and Peter Blanck. 2022. Workplace health surveillance and COVID-19: algorithmic health discrimination and cancer survivors. *Journal of Cancer Survivorship* 16, 1 (2022), 200–212.

[72] Gillian Z Heller, Maurizio Manuguerra, and Roberta Chow. 2016. How to analyze the Visual Analogue Scale: Myths, truths and clinical relevance. *Scandinavian journal of pain* 13, 1 (2016), 67–75.

[73] Alexander P Henkel, Stefano Bromuri, Deniz Iren, and Visara Urovi. 2020. Half human, half machine–augmenting service employees with AI for interpersonal emotion regulation. *Journal of Service Management* 31, 2 (2020), 247–265.

[74] Mireille Hildebrandt. 2014. *Location Data, Purpose Binding and Contextual Integrity: What's the Message?* Springer.

[75] Kasia Hitczenko, Henry R Cowan, Matthew Goldrick, and Vijay A Mittal. 2022. Racial and ethnic biases in computational approaches to psychopathology. , 285–288 pages.

[76] Arlie Russell Hochschild. 2022. The managed heart. In *Working In America*. Routledge, 40–48.

[77] Nicola Howe, Emma Giles, Dorothy Newbury-Birch, and Elaine McColl. 2018. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. *Journal of health services research & policy* 23, 2 (2018), 123–133.

[78] David C Howell. 2012. *Statistical methods for psychology.* Cengage Learning.

[79] Joop J Hox and Cora JM Maas. 2002. Sample sizes for multilevel modeling. (2002).

[80] Susan Jamieson. 2004. Likert scales: how to (ab)use them. *Medical Education* 38, 12 (Dec. 2004), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

[81] Guillermina Jasso. 2006. Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research* 34, 3 (2006), 334–423.

[82] Leslie K John, Alessandro Acquisti, and George Loewenstein. 2011. Strangers on a plane: Context-dependent willingness to divulge sensitive information. *Journal of consumer research* 37, 5 (2011), 858–873.

[83] Amba Kak. 2020. Regulating biometrics: Global approaches and urgent questions. *AI Now Institute, September* 1 (2020).

[84] Edward B Kang. 2023. On the Praxes and Politics of AI Speech Emotion Recognition. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 455–466.

[85] Nadia Karizat, Alexandra H Vinson, Shobita Parthasarathy, and Nazanin Andalibi. 2024. Patent applications as glimpses into the sociotechnical imaginary: Ethical speculation on the imagined futures of emotion AI for mental health monitoring and detection. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–43.

[86] Harmanpreet Kaur, Daniel McDuff, Alex C Williams, Jaime Teevan, and Shamsi T Iqbal. 2022. "I Didn't Know I Looked Angry": Characterizing Observed Emotion and Reported Affect at Work. In *CHI Conference on Human Factors in Computing Systems*. 1–18.

[87] Danielle Keats Citron. 2022. *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age.* W.W. Norton & Company.

[88] Flavius Kehr, Tobias Kowatsch, Daniel Wentzel, and Elgar Fleisch. 2015. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal* 25, 6 (2015), 607–635.

[89] Yeolib Kim, Boreum Choi, and Yoonhyuk Jung. 2018. Individual differences in online privacy concern. *Asia Pacific Journal of Information Systems* 28, 4 (2018), 274–289.

[90] Nikolaos Koutsouleris, Tobias U Hauser, Vasilisa Skvortsova, and Munmun De Choudhury. 2022. From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health* (2022).

[91] Brian Kropp. 2019. The Future of Employee Monitoring. //www.gartner.com/smarterwithgartner/the-future-of-employee-monitoring/

[92] Ponnurangam Kumaraguru and Lorrie Faith Cranor. 2005. Privacy indexes: a survey of Westin's studies. (2005).

[93] Jangho Kwon, Da-Hye Kim, Wanjoo Park, and Laehyun Kim. 2016. A wearable device for emotional recognition using facial expression and physiological response. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 5765–5768.

[94] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (nov 2018), 31 pages. https://doi.org/10.1145/3274371

[95] Hyunsoo Lee, Soowon Kang, and Uichin Lee. 2022. Understanding privacy risks and perceived benefits in open dataset collection for mobile affective computing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–26.

[96] Hwansoo Lee, Siew Fan Wong, Jungjoo Oh, and Younghoon Chang. 2019. Information privacy concerns and demographic characteristics: Data from a Korean media panel survey. *Government Information Quarterly* 36, 2 (2019), 294–303.

[97] Pedro Giovanni Leon, Justin Cranshaw, Lorrie Faith Cranor, Jim Graves, Manoj Hastak, Blase Ur, and Guzi Xu. 2012. What do online behavioral advertising privacy disclosures communicate to users?. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*. 19–30.

[98] Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. 2020. Privacy and activism in the transgender community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[99] Holly Lewis. 2016. *The politics of everybody: Feminism, queer theory, and marxism at the intersection.* Bloomsbury Publishing.

[100] Yunji Liang, Xiaolong Zheng, and Daniel D Zeng. 2019. A survey on big data-driven digital phenotyping of mental health. *Information Fusion* 52 (2019), 290–307.

[101] Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. 501–510.

[102] Steven G Luke. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior research methods* 49 (2017), 1494–1502.

[103] Naresh K Malhotra, Sung S Kim, and James Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.

[104] Peter Mantello, Manh-Tung Ho, Minh-Hoang Nguyen, and Quan-Hoang Vuong. 2021. Bosses without a heart: socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. *AI & society* (2021), 1–23.

[105] Ereni Markos, George R Milne, and James W Peltier. 2017. Information sensitivity and willingness to provide continua: a comparative privacy study of the United States and Brazil. *Journal of Public Policy & Marketing* 36, 1 (2017), 79–96.

[106] Kirsten Martin. 2015. Privacy notices as tabula rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online. *Journal of Public Policy & Marketing* 34, 2 (2015), 210–227.

[107] Kirsten Martin. 2018. The penalty for privacy violations: How privacy violations impact trust online. *Journal of Business Research* 82 (2018), 103–116.

[108] Kirsten Martin and Helen Nissenbaum. 2016. Measuring privacy: An empirical test using context to expose confounding variables. *Colum. Sci. & Tech. L. Rev.* 18 (2016), 176.

[109] Kirsten Martin and Helen Nissenbaum. 2020. What is it about location? *Berkeley Tech. LJ* 35 (2020), 251.

[110] Kirsten Martin and Katie Shilton. 2016. Why experience matters to privacy: How context-based experience moderates consumer privacy expectations for mobile applications. *Journal of the Association for Information Science and Technology* 67, 8 (2016), 1871–1882.

[111] Kelly D Martin and Patrick E Murphy. 2017. The role of data privacy in marketing. *Journal of the Academy of Marketing Science* 45 (2017), 135–155.

[112] Kirsten E Martin. 2012. Diminished or just different? A factorial vignette study of privacy as a social contract. *Journal of Business Ethics* 111, 4 (2012), 519–539.

[113] Alice E Marwick and Danah Boyd. 2018. Understanding privacy at the margins. *International Journal of Communication (19328036)* 12 (2018).

[114] Nora McDonald and Andrea Forte. 2020. The politics of privacy theories: Moving from norms to vulnerabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[115] Daniel McDuff, Eunice Jun, Kael Rowan, and Mary Czerwinski. 2019. Longitudinal Observational Evidence of the Impact of Emotion Regulation Strategies on Affective Expression. *IEEE Transactions on Affective Computing* 12, 3 (2019), 636–647.

[116] Derrick McIver, Mark L Lengnick-Hall, and Cynthia A Lengnick-Hall. 2018. A strategic approach to workforce analytics: Integrating science and agility. *Business Horizons* 61, 3 (2018), 397–407.

[117] Roisin McNaney, Catherine Morgan, Pranav Kulkarni, Julio Vega, Farnoosh Heidarivincheh, Ryan McConville, Alan Whone, Mickey Kim, Reuben Kirkham, and Ian Craddock. 2022. Exploring perceptions of cross-sectoral data sharing with people with Parkinson's. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[118] George R Milne, George Pettinico, Fatima M Hajjat, and Ereni Markos. 2017. Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. *Journal of Consumer Affairs* 51, 1 (2017), 133–161.

[119] Scott Monteith, Tasha Glenn, John Geddes, Peter C Whybrow, and Michael Bauer. 2022. Commercial Use of Emotion Artificial Intelligence (AI): Implications for Psychiatry. *Current Psychiatry Reports* (2022), 1–9.

[120] Sumitava Mukherjee, Jaison A Manjaly, and Maithilee Nargundkar. 2013. Money makes you reveal more: consequences of monetary cues on preferential disclosure of personal information. *Frontiers in psychology* 4 (2013), 839.

[121] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy expectations and preferences in an {IoT} world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 399–412.

[122] Jeff Nagy. 2022. Autism and the making of emotion AI: Disability as resource for surveillance capitalism. *New media & society* (2022), 14614448221109550.

[123] Yefim Natis and Jason Daigler. 2019. Gartner's Top Strategic Predictions for 2020 and Beyond: Technology Changes the Human Condition. https://www.gartner.com/document/3970846?ref=authrightrec&refval=3994528

[124] Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

[125] Helen Nissenbaum. 2011. A contextual approach to privacy online. *Daedalus* 140, 4 (2011), 32–48.

[126] Helen Nissenbaum. 2019. Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law* 20, 1 (2019), 221–256.

[127] Helen Nissenbaum. 2021. University of Washington's Department of Human Centered Design & Engineering 2021 Distinguished Lecture "Contextual Integrity: Breaking the Grip of Public-Private Distinction for Meaningful Privacy". https://www.youtube.com/watch?v=VPwmC0Sfe50. Accessed: 12/15/2023.

[128] Kobbi Nissim and Alexandra Wood. 2018. Is privacy privacy? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2128 (2018), 20170358.

[129] Steven L Nock and Thomas M Guterbock. 2010. Survey experiments. *Handbook of survey research* 2 (2010), 837–865.

[130] Leysan Nurgalieva, David O'Callaghan, and Gavin Doherty. 2020. Security and privacy of mHealth applications: a scoping review. *IEEE Access* 8 (2020), 104247–104268.

[131] Susan Moller Okin and Gender Justice. 1989. the Family. *New York* (1989).

[132] Judith S Olson, Jonathan Grudin, and Eric Horvitz. 2005. A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in computing systems*. 1985–1988.

[133] Yong Jin Park, Scott W Campbell, and Nojin Kwak. 2012. Affect, cognition and reward: Predictors of privacy protection online. *Computers in Human Behavior* 28, 3 (2012), 1019–1027.

[134] The European Parliament and Council of the European Union. [n. d.]. Art. 9 GDPR – Processing of special categories of personal data. https://gdpr-info.eu/art-9-gdpr/. Accessed: 2023-02-07.

[135] Vishal Patel, Austin Chesmore, Christopher M Legner, and Santosh Pandey. 2022. Trends in workplace wearable technologies and connected-worker solutions for next-generation occupational safety, health, and productivity. *Advanced Intelligent Systems* 4, 1 (2022), 2100099.

[136] José Pinheiro and Douglas Bates. 2006. *Mixed-effects models in S and S-PLUS*. Springer science & business media.

[137] Lee Rainie, Cary Funk, Monica Anderson, and Alec Tyson. 2022. AI and Human Enhancement: Americans' Openness is Tempered by a Range of Concerns. March 2022 (March 2022), 164.

[138] Manuel Rausch and Michael Zehetleitner. 2014. A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consciousness and cognition* 28 (2014), 126–140.

[139] Lauren Rhue. 2018. Racial influence on automated perceptions of emotions. *Available at SSRN 3281765* (2018).

[140] Neil Richards. 2022. *Why Privacy Matters*. Oxford University Press.

[141] Clara E Rodriguez, Michael H Miyawaki, and Grigoris Argeros. 2013. Latino racial reporting in the US: To be or not to be. *Sociology Compass* 7, 5 (2013), 390–403.

[142] Kat Roemmich and Nazanin Andalibi. 2021. Data Subjects' Conceptualizations of and Attitudes Toward Automatic Emotion Recognition-Enabled Wellbeing Interventions on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–34.

[143] Kat Roemmich, Shanley Corvite, Cassidy Pyle, Nadia Karizat, and Nazanin Andalibi. 2024. Emotion AI Use in US Mental Healthcare: Potentially Unjust and Techno-Solutionist. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–46.

[144] Kat Roemmich, Tillie Rosenberg, Serena Fan, and Nazanin Andalibi. 2023. Values in Emotion Artificial Intelligence Hiring Services: Technosolutions to Organizational Problems. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–28.

[145] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[146] Andrew J Rohm and George R Milne. 2004. Just what the doctor ordered: The role of information sensitivity and trust in reducing medical information privacy concern. *Journal of Business Research* 57, 9 (2004), 1000–1011.

[147] Samiha Samrose, Wenyi Chu, Carolina He, Yuebai Gao, Syeda Sarah Shahrin, Zhen Bai, and Mohammed Ehsan Hoque. 2019. Visual cues for disrespectful conversation analysis. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 580–586.

[148] Samiha Samrose, Wenyi Chu, Carolina He, Yuebai Gao, Syeda Sarah Shahrin, Zhen Bai, and Mohammed Ehsan Hoque. 2019. Visual Cues for Disrespectful Conversation Analysis. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 580–586. https://doi.org/10.1109/ACII.2019.8925440 ISSN: 2156-8111.

[149] Franklin E Satterthwaite. 1941. Synthesis of variance. *Psychometrika* 6, 5 (1941), 309–316.

[150] Morgan Klaus Scheuerman, Stacy M Branham, and Foad Hamidi. 2018. Safe spaces and safe places: Unpacking technology-mediated experiences of safety and harm with transgender people. *Proceedings of the ACM on Human-computer Interaction* 2, CSCW (2018), 1–27.

[151] Dagmar Schuller and Björn W Schuller. 2018. The age of artificial emotional intelligence. *Computer* 51, 9 (2018), 38–46.

[152] Kim Bartel Sheehan. 2002. Toward a typology of Internet users and online privacy concerns. *The information society* 18, 1 (2002), 21–32.

[153] Nelson Shen, Lydia Sequeira, Michelle Pannor Silver, Abigail Carter-Langford, John Strauss, and David Wiljer. 2019. Patient privacy perspectives on health information exchange in a mental health context: qualitative study. *JMIR mental health* 6, 11 (2019), e13306.

[154] Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979), 420.

[155] Scott Skinner-Thompson. 2020. *Privacy at the Margins*. Cambridge University Press.

[156] P Slovic. 2000. The perception of risk. Earthscan. *London, UK* (2000).

[157] H Jeff Smith, Sandra J Milberg, and Sandra J Burke. 1996. Information privacy: Measuring individuals' concerns about organizational practices. *MIS quarterly* (1996), 167–196.

[158] Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih-Fu Chang, and Maja Pantic. 2017. A survey of multimodal sentiment analysis. *Image and Vision Computing* 65 (2017), 3–14.
[159] Daniel J Solove. 2005. A taxonomy of privacy. *U. Pa. l. Rev.* 154 (2005), 477.
[160] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.
[161] Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (2019), 62–65.
[162] Luke Stark and Jesse Hoey. 2021. The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 782–793.
[163] Luke Stark, Amanda Stanhaus, and Denise L Anthony. 2020. "i don't want someone to watch me while i'm working": Gendered views of facial recognition technology in workplace surveillance. *Journal of the Association for Information Science and Technology* 71, 9 (2020), 1074–1088.
[164] Jenny Tang, Eleanor Birrell, and Ada Lerner. 2022. Replication: How well do my results generalize now? The external validity of online privacy and security surveys. In *Eighteenth symposium on usable privacy and security (SOUPS 2022).* 367–385.
[165] The American Journal of Managed Care. 2006. Vulnerable Populations: Who Are They? *The American Journal of Managed Care (AJMC)* 12, 13 (Nov. 2006), 348–352. https://www.ajmc.com/view/nov06-2390ps348-s352 Publisher: MJH Life Sciences.
[166] Laura K. Thompson, Margaret M. Sugg, and Jennifer R. Runkle. 2018. Adolescents in crisis: A geographic exploration of help-seeking behavior using data from Crisis Text Line. *Social Science & Medicine* 215 (Oct. 2018), 69–79. https://doi.org/10.1016/j.socscimed.2018.08.025
[167] Somchanok Tivatansakul, Michiko Ohkura, Supadchaya Puangpontip, and Tiranee Achalakul. 2014. Emotional healthcare system: Emotion detection by facial expressions using Japanese database. In *2014 6th computer science and electronic engineering conference (CEEC).* IEEE, 41–46.
[168] Jan Tolsdorf and Florian Dehling. 2020. In our employer we trust: mental models of office workers' privacy perceptions. In *International Conference on Financial Cryptography and Data Security.* Springer, 122–136.
[169] Kate E Toth and Carolyn S Dewa. 2014. Employee decision-making about disclosure of a mental disorder at work. *Journal of occupational rehabilitation* 24 (2014), 732–746.
[170] Horst Treiblmaier and Peter Filzmoser. 2011. Benefits from Using Continuous Rating Scales in Online Survey Research.. In *ICIS.*
[171] Thomas Tsiampalis and Demosthenes B Panagiotakos. 2020. Missing-data analysis: socio-demographic, clinical and lifestyle determinants of low response rate on self-reported psychological and nutrition related multi-item instruments in the context of the ATTICA epidemiological study. *BMC Medical Research Methodology* 20, 1 (2020), 1–13.
[172] Mary P Tully, Kyle Bozentko, Sarah Clement, Amanda Hunn, Lamiece Hassan, Ruth Norris, Malcolm Oswald, and Niels Peek. 2018. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. *Journal of medical Internet research* 20, 3 (2018), e7763.
[173] U.S. Department of Health and Human Services. 2018. Code of Federal Regulations, Title 45: Public Welfare, Part 46: Protection of Human Subjects, Section 46.104: Exempt Research. https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46#46.104. Revised Common Rule.
[174] Eric M Uslaner. 2004. Trust, civic engagement, and the Internet. *Political Communication,* 21, 2 (2004), 223–242.
[175] André Calero Valdez and Martina Ziefle. 2019. The users' perspective on the privacy-utility trade-offs in health recommender systems. *International Journal of Human-Computer Studies* 121 (2019), 108–121.
[176] Madison Van Oort. 2019. The emotional labor of surveillance: Digital control in fast fashion retail. *Critical Sociology* 45, 7-8 (2019), 1167–1179.
[177] Jukka Varelius. 2006. The value of autonomy in medical ethics. *Medicine, Health Care and Philosophy* 9 (2006), 377–388.
[178] William N Venables and Brian D Ripley. 2002. Random and mixed effects. In *Modern applied statistics with S.* Springer, 271–300.
[179] Rosalie Waelen and Michał Wieczorek. 2022. The Struggle for AI's Recognition: Understanding the Normative Implications of Gender Bias in AI with Honneth's Theory of Recognition. *Philosophy & Technology* 35, 2 (June 2022), 53. https://doi.org/10.1007/s13347-022-00548-w
[180] Ari Ezra Waldman. 2018. *Privacy as trust: Information privacy for an information age.* Cambridge University Press.
[181] Michael Walzer. 1984. Spheres of justice: a defense of pluralism and equality.
[182] Suzanne P Weisband and Bruce A Reinig. 1995. Managing user perceptions of email privacy. *Commun. ACM* 38, 12 (1995), 40–47.
[183] Alan F Westin and Danielle Maurici. 1998. *E-commerce & privacy: What net users want.* Privacy & American Business Hackensack, NJ.
[184] Stephen Wood, Valerio Ghezzi, Claudio Barbaranelli, Cristina Di Tecco, Roberta Fida, Maria Luisa Farnese, Matteo Ronchetti, and Sergio Iavicoli. 2019. Assessing the risk of stress in organizations: getting the measure of organizational-level stressors. *Frontiers in Psychology* 10 (2019), 2776.
[185] James D Wright, Peter V Marsden, et al. 2010. Survey research and social science: History, current practice, and future prospects. *Handbook of survey research* (2010), 3–26.
[186] Heng Xu, Hao Wang, and Hock-Hai Teo. 2005. Predicting the usage of P2P sharing software: The role of trust and perceived risk. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences.* IEEE, 201a–201a.
[187] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. 2020. Investigating bias and fairness in facial expression recognition. In *European Conference on Computer Vision.* Springer, 506–523.
[188] Biqiao Zhang and Emily Mower Provost. 2019. Automatic recognition of self-reported and perceived emotions. In *Multimodal Behavior Analysis in the Wild.* Elsevier, 443–470.
[189] Dongsong Zhang, Jaewan Lim, Lina Zhou, and Alicia A Dahl. 2021. Breaking the Data Value-Privacy Paradox in Mobile Mental Health Systems Through User-Centered Privacy Protection: A Web-Based Survey Study. *JMIR Mental Health* 8, 12 (2021), e31633.

[190] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. 2020. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion* 59 (2020), 103–126.

[191] Shikun Zhang, Yuanyuan Feng, and Norman Sadeh. 2021. Facial Recognition: Understanding Privacy Concerns and Attitudes Across Increasingly Diverse Deployment Scenarios. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, 243–262. https://www.usenix.org/conference/soups2021/presentation/zhang-shikun

[192] Kathryn Zickuhr. 2021. Workplace Surveillance Is Becoming the New Normal for US Workers. *Washington Center for Equitable Growth. https://equitablegrowth. org/research-paper/workplace-surveillance-is-becomingthe-new-normal-for-us-workers/. Institute for Research on Labor and Employment University of California, Berkeley* 2521 (2021), 94720–5555.

7 APPENDICES

A PLOTTED COEFFICIENTS WITH ERROR BARS

B POST-TEST SOCIO-DEMOGRAPHIC QUESTIONS

1. Please indicate your current employment status. Select all that apply.

- Employed Full-Time
- Employed Part-Time
- Looking for work
- Not in the paid workforce (retired, full-time caregiving, full-time student, etc.)
- Other

2, What is the highest level of school you have completed or the highest degree you have received?

- No formal schooling
- Some grade school
- High school graduate (high school diploma or equivalent including GED)
- Some college
- Technical, vocational, or trade school
- Associate degree in college (2-year)
- Bachelor's degree in college (4-year)
- Master's degree
- Professional degree (JD, MD)
- Doctoral degree

3. What is your year of birth? <text box>

4. Please describe your race/ethnicity. Select all that apply.

- African
- African-American or Black
- Asian-American
- East Asian
- Hispanic or Latino/a/x
- Indigenous American or First Nations
- Middle Eastern
- South Asian
- Southeast Asian
- White

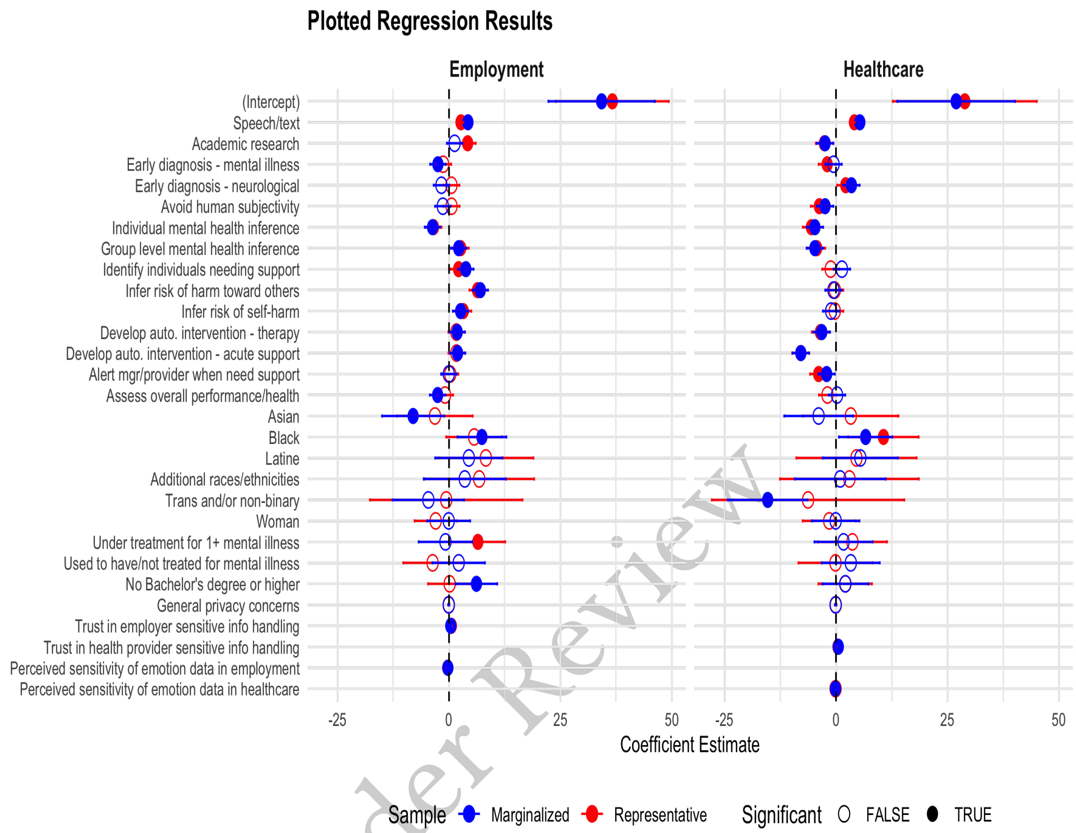

Fig. 8. Coefficient Plot with Error Bars Each point represents the tested independent variable; its position on the x-axis indicates the estimated effect size on reported comfort. Filled circles signify statistically significant relationships; open circles represent non-significant relationships; the color red represents estimated negative relationships; the color blue represents estimated positive relationships. Vertical dashed lines mark the zero line. Error bars display 95% confidence intervals around coefficient estimates. The plot offers insights into the direction, significance, and uncertainty of variable effects.

- Not listed, please specify <text box>
- Prefer not to answer

5. Please describe your mental health status. Select all that apply.

- I have a mental health condition and it has not been formally diagnosed
- I have a mental health condition that has been formally diagnosed
- I am being treated for a mental health condition, and that treatment includes medication
- I am being treated for a mental health condition, not with medication
- I do not have a mental health condition

- I used to have a mental health condition and I no longer do
- I have multiple mental health conditions. Some are diagnosed, some are not
- I have multiple mental health conditions. I take medication for some, and do not for others

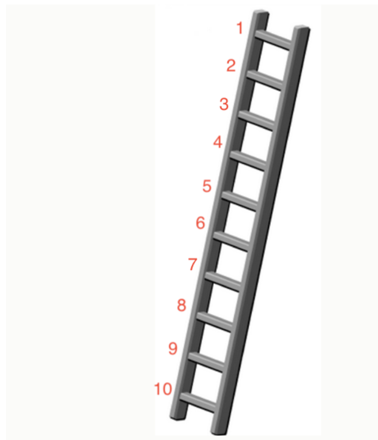

Fig. 9. MacArthur Scale of Subjective Social Status)

6. At the top of the ladder are the people who are the best off, those who have the most money, most education, and best jobs. At the bottom are the people who are the worst off, those who have the least money, least education, worst jobs, or no job. Select the number next to the rung that best represents where you think you stand on the ladder.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- Prefer not to answer

C POST-TEST INDIVIDUAL BELIEF QUESTIONS

C.1 General Privacy Concerns

Rate your agreement, from 0= "strongly disagree" to 100 = "strongly agree," with the following:

- All things considered, the internet causes serious privacy problems.
- Compared to others, I am more sensitive about the way my personal information is handled.
- To me, it is the most important thing to keep my privacy intact from companies and institutions.
- I believe other people are too much concerned with online privacy issues.

- Compared with other subjects on my mind, personal privacy is very important.
- I am concerned about threats to my personal privacy today.

C.2 Risk Beliefs Regarding Employers' and Healthcare Providers' Handling of Sensitive Information

Rate your agreement, from 0= "strongly disagree" to 100 = "strongly agree," with the following:

- In general, it is risky to give sensitive information to **employers**.
- In general, it is risky to give sensitive information to **healthcare providers**.
- There is a high potential for loss associated with **employers** handling sensitive data about me.
- There is a high potential for loss associated with **healthcare providers** handling sensitive data about me.
- There is too much uncertainty associated with giving sensitive information to **employers**.
- There is too much uncertainty associated with giving sensitive information to **healthcare providers**.
- Providing **employers** with sensitive information would involve many unexpected problems.
- Providing **healthcare providers** with sensitive information would involve many unexpected problems.
- I feel safe giving sensitive information to **employers**.
- I feel safe giving sensitive information to **healthcare providers**.

C.3 Trust Beliefs Regarding Employers' and Healthcare Providers' Handling of Sensitive Information

Rate your agreement, from 0= "strongly disagree" to 100 = "strongly agree," with the following:

- **Employers** are trustworthy in handling sensitive information about me.
- **Healthcare providers** are trustworthy in handling sensitive information about me.
- **Employers** would tell the truth and fulfill promises related to how they use sensitive information about me.
- **Healthcare providers** would tell the truth and fulfill promises related to how they use sensitive information about me.
- I trust that **employers** would keep my best interests in mind when dealing with sensitive information about me.
- I trust that **healthcare providers** would keep my best interests in mind when dealing with sensitive information about me.
- **Employers** are in general predictable and consistent regarding the usage of **employees'** sensitive information.
- **Healthcare providers** are in general predictable and consistent regarding the usage of **patients'** sensitive information.
- **Employers** are always honest with **employees** when it comes to using their sensitive information about **employees**.
- **Healthcare providers** are always honest with **patients** when it comes to using their sensitive information about **patients**.

C.4 Perceived Sensitivity of Emotional Information and Other Sensitive Data Types when Handled by Employers and Healthcare Providers

Rate your agreement, from 0= "strongly disagree" to 100 = "strongly agree," with the following:

- When an **employer** has access to information about your **emotional states** (states of feeling like emotion or mood, including but not limited to stress, anxiety, depression, boredom, calm, fear, fatigue, attentiveness, happiness, sadness, disgust, surprise, and/or anger), how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **emotional states** (states of feeling like emotion or mood, including but not limited to stress, anxiety, depression, boredom, calm, fear, fatigue, attentiveness, happiness, sadness, disgust, surprise, and/or anger), how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **political opinions**, how SENSITIVE do you consider this information?
- When an **healthcare provider** has access to information about your **political opinions**, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **religious beliefs**, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **religious beliefs**, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **biometric data**, such as your fingerprints, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **biometric data**, such as your fingerprints, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **health**, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **health**, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **sex life or sexual orientation**, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **sex life or sexual orientation**, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **genetic information**, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **genetic information**, how SENSITIVE do you consider this information?
- When an **employer** has access to information about your **current or past union membership**, how SENSITIVE do you consider this information?
- When a **healthcare provider** has access to information about your **current or past union membership**, how SENSITIVE do you consider this information?

Received 20 December 2023; revised 17 May 2025