Article:

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Commentary

"Empathy Code": The Dangers of Automating Empathy in Business

$\mathbb{S}$ Sage

Nicola Thomas[1] (iD) and Niall Docherty[1]

Abstract

Organizations are increasingly adopting "Emotional AI" to monitor and influence employee emotions, aiming to create more empathetic workplaces. However, we argue that automating empathy risks fostering *empathy skepticism*, alienating employees, exacerbating mental health issues, and eroding trust. We call on organizations to address the root causes of negative workplace emotions and leverage AI as a tool to complement—rather than replace—empathy, fostering workplaces that genuinely prioritize care and trust.

Keywords

artificial intelligence, emotional AI, emotions, empathy

Employee mental health has reached an all-time low, with a third of workers experiencing emotional exhaustion (American Psychological Association, 2023). In response, companies have turned to "Emotional AI"—tools that track voice, use of language, facial expressions, and heart rate to identify signs of distress—claiming these technologies can embed higher levels of empathy in organizations and ultimately drive efficiency and productivity. Emotional AI is portrayed as a techno-fix for the workplace mental health crisis.

[1]University of Sheffield, UK

Corresponding Author:
Nicola Thomas, Institute of Work Psychology, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.
Email: Nicola.thomas@sheffield.ac.uk

However, automating empathy overlooks the complex social, cultural, and political factors contributing to deteriorating employee mental health. Framed as a techno-solution, Emotional AI is a reductive (if well-meaning) intervention that fails to understand the root causes of the issues it seeks to address, all while claiming to increase worker outputs. Against this backdrop, we further argue that automating empathy poses a significant threat to individuals, businesses, and society for two core reasons. First, it dehumanizes both employees and employers, reducing humans to computers upon which *empathy code* is run. Second, it results in *empathy skepticism*—a mistrust of empathetic displays at work—eroding trust in human connection.

The Dehumanization of Empathy, Introducing *Empathy Code*

Various commercial Emotional AI products are on the market, such as chatbots designed to ease employee stress and AI facial expression monitoring that prompts employees to smile. By using Emotional AI, organizations treat empathy as an instrumental activity, outsourcing emotional processes to technology, and positioning empathy as a process that can be scripted and run like a computer code. However, the premise that an algorithm can recreate human empathy fundamentally misunderstands what empathy is. Empathy cannot be separated from the humans *feeling* it. Beyond understanding someone's hardship cognitively, empathy elicits a genuine emotional response of concern or compassion for another person's struggles, driving consistently better workplace outcomes including stronger, more supportive relationships, and lower burnout (Longmire & Harrison, 2018). This makes empathy uniquely *target-centric*, focusing almost entirely on the feelings of the other person, the "target" of empathy.

Imagine being in a difficult situation where, instead of speaking to your manager, an algorithm detects your negative emotions through your heart rate and facial expressions, and encourages you to talk with an "empathetic AI chatbot." Taking the place of an active listener, such technologies track your emotional well-being by examining the language you employ during text-based (or verbal) conversations. When the bot identifies your distress, it is trained to respond with an "empathetic" tone and proffer solutions, all through a sophisticated linguistic imitation akin to how empathy is usually expressed between humans. Yet imitating human speech alone is meaningless. Think of the way parrots could mimic empathetic utterances. A parrot may be able to say the right things but consider how strange and unconvincing its words of support would feel in times of need. Think of the same for bots.

Troublingly, AI evangelists chalk these concerns up to outdated thinking. Rather than viewing the interpersonal qualities of human existence as unique, some claim that exponential advances in computational power have shown us

that humans, at our core, are computational. The generative capacity of artificial intelligence—its neural networks and machine learning—is presented as evidence that what we once thought of as distinctly human is replicable by machines. Such beliefs are known as computationalism, a view that human minds are computers (Golumbia, 2009). By viewing human cognitive abilities as computers, computationalism claims that the "human and social experience can be explained [and replicated] via computational processes" (Golumbia, 2009, p. 8). Taken to its extreme, using Emotional AI to automate empathy in the workplace signals a continuation of this type of thought, divorcing human emotions and feelings from the individuals experiencing them.

Instead, empathy is broken down into words and phrases, like a kind of empathy code made up of ones and zeros. In this framework, driven by capitalist logic, even empathy—a deeply human quality—can be automated and scaled for profit. Human emotion becomes data, another resource to be measured and managed; another workplace function to maximize efficiency. Such an approach does not elevate employee well-being, rather, it undermines it, reducing empathy to an instrument of control. By using Emotional AI, truly caring for employees is pushed aside. The negative impact this could have on relationships, mental health, employee morale, and sense of cohesion at work is clear.

The Threat of Empathy Skepticism

As Emotional AI becomes more seamlessly integrated into the workplace, the boundary between human and AI-driven empathy becomes blurred. Companies now use Emotional AI email add-ons that embed pre-scripted empathetic phrases, for example, meaning employees may struggle to tell whether expressions of concern come from a manager's genuine empathy or from an algorithm programmed to mimic care.

Once-trustworthy messages now risk becoming suspect, leading employees to view expressions of empathy as insincere or even manipulative; a product of AI rather than genuine human concern. Instead of supporting well-being, Emotional AI may backfire and create a perception of hypocrisy that erodes trust and harms mental health, generating skepticism. Research supports this concern: studies show that when people realize empathetic messages are AI-generated, their sense of being understood diminishes, often replaced by feelings of reduced authenticity and insincerity (Yin et al., 2024).

Employees could interpret empathetic gestures made by AI as reminders that their emotions are being commodified. Or worse, this technology could make employees feel constantly surveilled (Chhillar & Aguilera, 2022). Over time, this distrust could create a toxic work environment, worsening employee mental health and undermining the supportive relationships that contribute to

a healthy workplace. Overall, we argue that the empathy code fundamentally misinterprets how empathy is experienced, curtails the formation of emotional bonds in the workplace, and, perhaps most importantly, devalues the well-being of human colleagues.

A Call to Action: Reclaiming Authentic Empathy

The true power of empathy lies in its ability to inspire genuine actions and emotions in response to complex cues in human interactions. We call on organizations and researchers to critically (re)assess the implementation of Emotional AI, and take the following steps to safeguard trust and well-being in workplaces:

1. *Address Root Causes of Negative Emotion*: Rather than seeking quick technological fixes for stress, focus on systemic contributing factors, such as excessive workloads, poor management practices, and toxic work cultures. Implement policies to promote a psychologically safe work environment in response.
2. *Use AI to Enhance, Not Replace, Human Empathy*: Leverage AI for routine administrative tasks, freeing managers' time and capacity to provide direct, authentic emotional support to employees. Organizations should avoid relying on Emotional AI for tasks requiring emotional nuance.
3. *Focus on Organization Culture*: Foster a workplace culture that values authenticity, transparency, and genuine care. Build systems and practices that prioritize interpersonal relationships, ensuring employees feel supported and valued.
4. *Ensure Transparency and Ethical Use*: Clearly disclose when Emotional AI tools are in use and provide employees with the option to opt out of these systems.
5. *Collaborate with Policymakers and Researchers*: Engage with regulatory frameworks such as the EU AI Act, specifically understanding the implications of Emotional AI being classified as a high-risk AI system. Organizations must ensure compliance by conducting regular audits, ensuring transparent algorithms, and establishing safeguards to protect employees' emotional data.

By implementing these actions, organizations can create workplaces that address mental health challenges at their core. Rather than dehumanizing the workplace with superficial technological fixes such as Emotional AI, organizations can lead with empathy, demonstrating that genuine care is not just an ethical imperative but a strategic advantage.

Declaration of Conflicting Interests

Funding

ORCID iD

Nicola Thomas https://orcid.org/0000-0003-3329-6533

References

American Psychological Association. (2023). *Work in America: Workplaces as engines of psychological health and well-being.* https://www.apa.org/pubs/reports/work-in-america/2023-workplace-health-well-being

Chhillar, D., & Aguilera, R. V. (2022). An eye for artificial intelligence: Insights into the governance of artificial intelligence and vision for future research. *Business & Society*, *61*(5), 1197–1241.

Golumbia, D. (2009). *The cultural logic of computation*. Harvard University Press.

Longmire, N. H., & Harrison, D. A. (2018). Seeing their side versus feeling their pain: Differential consequences of perspective-taking and empathy at work. *Journal of Applied Psychology*, *103*(8), 894–915.

Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences*, *121*(14), e2319112121.

Author Biographies

Nicola Thomas, PhD, is a lecturer in work psychology at the Institute of Work Psychology, University of Sheffield. Her research explores workplace emotions, with a particular focus on AI and gender dynamics. A dedicated science communicator, Nicola frequently appears on BBC platforms to make science accessible to the public.

Niall Docherty, PhD, is a lecturer in data, AI and society in the Information School at The University of Sheffield. His research examines the philosophical and political stakes of digital well-being. His first book *Healthy Users* (UC Press) explores these themes, and his work has appeared in interdisciplinary journals such as *ACM CHI, Information, Communication* and *Society, Science as Culture*, and *Social Media + Society.*