

The Risks of Using AI to Interpret Human Emotions

 hbr.org/2019/11/the-risks-of-using-ai-to-interpret-human-emotions

Mark Purdy John Zealley Omaro Maseli

November 18,
2019

Loading...

Executive Summary

Companies have historically used focus groups and surveys to understand how people felt. Now, emotional AI technology can help businesses capture the emotional reactions of both employees and consumers in real time — by decoding facial expressions, analyzing voice patterns, monitoring eye movements, and measuring neurological immersion levels, for example. The ultimate outcome is a much better understanding both of workers and customers. But, because of the subjective nature of emotions, emotional AI is especially prone to bias. AI is often also not sophisticated enough to understand cultural differences in expressing and reading emotions, making it harder to draw accurate conclusions. For instance, a smile might mean one thing in Germany and another in Japan. Confusing these meanings can lead businesses to make wrong decisions. Imagine a Japanese tourist needing assistance while visiting a shop in Berlin. Using emotion recognition to prioritize which customers to support, the shop assistant might mistake their smile — a sign of politeness back home — as an indication that they don't require help. If left unaddressed, conscious or unconscious emotional biases like this can perpetuate stereotypes and assumptions at an unprecedented scale.



Don Farrall/Getty Images

What do people really feel?

This has never been an easy thing for companies to determine. For one thing, emotions are inherently difficult to read. For another, there's often a disconnect between what people *say* they feel and what they *actually* feel.

Consider how people respond to Super Bowl commercials. In 2018, TV viewers voted Amazon's "[Alexa Loses Her Voice](#)" — where celebrities attempt to (unsuccessfully) replace Alexa — as the best commercial, according to the [USA Today Ad Meter](#). Diet Coke's "[Groove](#)," which featured a woman dancing awkwardly after drinking a can of Diet Coke Twisted Mango, was rated as the worst commercial. Based on this poll, one might conclude that the Alexa commercial had the bigger impact. Not so, according to Paul Zak, neuroscience researcher and chief executive officer of [Immersion Neuroscience](#), whose team [studied](#) people's neurologic immersion in the ads. Zak's team assessed viewers' level of emotional engagement by measuring changes in oxytocin levels, the brain's "neural signature of emotional resonance." The research found that "Groove" actually had greater impact — proof to Zak that for Super Bowl commercials, there is "zero correlation" between what people say and how they subconsciously feel.

When we interviewed Zak about this phenomenon, he summed it up by saying: "People lie, their brains don't."

A lot of companies use focus groups and surveys to understand how people feel. Now, emotional AI technology can help businesses capture the emotional reactions in real time — by decoding facial expressions, analyzing voice patterns, monitoring eye movements, and

measuring neurological immersion levels, for example. The ultimate outcome is a much better understanding of their customers — and even their employees.

The Risks of Bias in Emotional AI

Because of the subjective nature of emotions, emotional AI is especially prone to bias. For example, [one study](#) found that emotional analysis technology assigns more negative emotions to people of certain ethnicities than to others. Consider the ramifications in the workplace, where an algorithm consistently identifying an individual as exhibiting negative emotions might affect career progression.

Insight Center

AI and Bias

Sponsored by SAS

Building fair and equitable machine learning systems.

AI is often also not sophisticated enough to understand cultural differences in expressing and reading emotions, making it harder to draw accurate conclusions. For instance, a smile might mean one thing in Germany and another in Japan. Confusing these meanings can lead businesses to make wrong decisions. Imagine a Japanese tourist needing assistance while visiting a shop in Berlin. If the shop used emotion recognition to prioritize which customers to support, the shop assistant might mistake their smile — a sign of politeness back home — as an indication that they didn't require help.

In short, if left unaddressed, conscious or unconscious emotional bias can perpetuate stereotypes and assumptions at an unprecedented scale.

How Businesses Can Prevent Bias from Seeping Into Common Use Cases

Based on our research and experience working with global clients, we see businesses using emotional AI technology in four ways. Through each, the implications of algorithmic bias are a clear reminder that business and technology leaders must understand and prevent such biases from seeping in.

Understanding how emotionally engaged employees actually are. When AI is used to gauge employee emotions, it can have serious impacts on how work is allocated. For example, employees often think they're in the right role, but upon trying new projects might find their skills are better aligned elsewhere. [Some companies](#) are already allowing employees to try different roles once a month to see what jobs they like most. Here's where bias in AI could reinforce existing stereotypes. For example, in the U.S., where [89% of civil engineers and 81% of first-line police and detective supervisors are male](#), an algorithm that

has been conditioned to analyze male features might struggle to read emotional responses and engagement levels among female recruits. This could lead to flawed role allocation and training decisions.

Improving the ability to create products that adapt to consumer emotions. With emotion tracking, product developers can learn which features elicit the most excitement and engagement in users. Take, for example, [Affectiva's Auto AI platform](#), which can recognize emotions like joy and anger and adapt a vehicle's in-cabin environment accordingly. Cameras and microphones can pick up on passenger drowsiness — and may lower the temperature or jolt the seatbelt as a result. A smart assistant might change its tone in response to a frustrated passenger. With emotional AI, any product or service — whether in the car or elsewhere — can become an adaptive experience. But a biased adaptive in-cabin environment could mean that some passengers are misunderstood. Elderly people, for example, might be more likely to be wrongly identified as having driver fatigue (the older the age of the face, the less likely it is that expressions are accurately decoded). And as these systems become more commonplace, insurance companies are going to want a piece of the data. This could mean higher premiums for older people, as the data would suggest that, despite many prompts to rest, the driver pressed on.

Improving tools to measure customer satisfaction. Companies like Boston-based startup [Cogito](#) are giving businesses the tools to help their employees interact better with customers. Its algorithms can not only identify “compassion fatigue” in customer service agents, but can also guide agents on how to respond to callers via an app. An upset customer might, for example, call to complain about a product. Recording and analyzing the conversation, Cogito's platform would then suggest that the agent slow down or prompt them on when to display empathy. A biased algorithm, perhaps skewed by an accent or a deeper voice, might result in some customers being treated better than others — pushing those bearing the brunt of bad treatment away from the brand. A male caller could be subject to less empathy than a woman, reinforcing societal perceptions of men as “emotionally strong.” On the flip side, a female caller may be viewed as a less tough negotiator, resulting in less compensation being offered. Ironically, the agents themselves may not even possess these biases, but clouded by the misconception that the algorithms are highly accurate, they may follow their advice blindly. In this way, biases spread, unquestioned and systematically.

Transforming the learning experience. Emotional insights could be used to augment the learning experience across all ages. It could, for example, allow teachers to design lessons that spur maximum engagement, putting key information at engagement peaks and switching content at troughs. It also offers insights into the students themselves, helping to identify who needs more attention. [China](#) is already introducing emotion detection systems into classrooms to track how focused students are. But, if biases exist, wrongly suggesting someone is disengaged could result in learning experiences tailored toward certain groups

rather than others. Think about different learning styles: Some people are visual learners. Some learn by doing. Others favor intense solitary concentration. But an algorithm, perhaps designed by a visual learner, might completely miss or misinterpret such cues. Incorrect engagement readings could affect learning outcomes all the way to the workplace, meaning that even in work training programs, only a fraction of employees can enjoy full professional development. Such misassumptions could affect learning outcomes all the way to the workplace, meaning that even in work training programs, only a fraction of employees can enjoy full professional development.

Avoiding Bias in AI

As more and more companies incorporate emotional AI in their operations and products, it's going to be imperative that they're aware of the potential for bias to creep in and that they actively work to prevent it.

Whether it is the subjective nature of emotions, or discrepancies in emotions, it is clear that detecting emotions is no easy task. Some technologies are better than others at tracking certain emotions, so combining these technologies could help to mitigate bias. In fact, a Nielsen study testing the accuracy of neuroscience technologies such as facial coding, biometrics, and electroencephalography (EEG) found that when used alone, accuracy levels were at 9%, 27%, and 62% respectively. When combined, accuracy levels shot up to 77%. Testing the results with a survey brought this up to 84%. Such combinations therefore serve as a check on the accuracy of results — a referencing system of sorts.

But accounting for cultural nuances in algorithms will take more than just combining and referencing multiple technologies. Having diverse teams creating emotional AI algorithms will be crucial to keeping bias at bay and fully capturing the complexity of emotions. This means not just gender and ethnic diversity, but also diversity in socioeconomic status and views – negating anything from xenophobia to homophobia to ageism. The more diverse the inputs and data points, the more likely it is that we'll be able to develop AI that's fair and unbiased.

Companies will also need to be vigilant about not perpetuating historical biases when training emotional AI. While historical data might be used as a basis to train AI on different emotional states, real-time, live data will be needed for context. Take smiles, for example. One study showed that of the 19 different types of smile, only six happen when people are having a good time. We also smile when we are in pain, embarrassed, and uncomfortable — distinctions that can only be drawn with context.

In sum, emotional AI will be a powerful tool indeed, forcing businesses to reconsider their relationships with consumers and employees alike. It will not only offer new metrics to understand people, but will also redefine products as we know them. But as businesses

foray into the world of emotional intelligence, the need to prevent biases from seeping in will be essential. Failure to act will leave certain groups systematically more misunderstood than ever — a far cry from the promises offered by emotional AI.

The authors would like to thank Accenture Research colleagues Xiao Chang, Paul Barbagallo, Dave Light, and H. James Wilson for their significant contributions to this article.