Basic emotions and AI

Author information removed for blind review

The primary discipline of this contribution is philosophy. This submission is intended as a paper.

ABSTRACT:

I contend that the widespread use of Paul Ekman's theory of basic emotions in artificial intelligence is scientifically and philosophically questionable, despite its apparent practical utility. In this paper, I briefly outline the key assumptions of Ekman's theory and selected critiques of emotional essentialism as contextual background, before focusing primarily on whether the revised position preserves any explanatory or empirical value in light of sustained criticism.

Researchers still cannot agree on a specific definition of emotion, but many classifications of this psychological phenomenon have been developed. One of them, also known as the "basic emotion theory", assumes that:

1) There are certain universal basic emotions in every human being
2) These can be identified in an individual through the analysis of facial expressions and microexpressions.

The author of this concept is Paul Ekman, whose research findings, despite criticism from anthropologists, psychologists, neuroscientists, and philosophers, have become firmly anchored in much of contemporary reality: national security systems at airports, education, hiring start-ups, systems that purport to detect psychiatric illness and policing programmes that claim to predict violence.
Existing research indicates that there are flaws in his model, which they assume to be overly simplistic, as many reactions and descriptions of emotional experiences do not fit into these categories (Coppini et al., 2023). Furthermore, Barrett et al. (2019) argue that there is no consistent evidence for a relationship between facial configuration and a specific emotional state, and that context, experience, and the observer's perceptual abilities influence the interpretation of a given emotion.

Based on Ekman's research, emotion recognition was developed, which aims to identify and interpret human emotions and mental states by analysing various cues, such as facial expressions, body language, and speech patterns.

Nowadays it is mostly a subfield of AI that by leveraging machine learning algorithms and computer vision techniques, it attempts to classify emotions into categories such as happiness, sadness, anger, fear, surprise, and disgust (Crawford, 2021).
It can be found in programs and technologies like Affectiva - a software company that has developed an emotion recognition platform using computer vision and deep learning algorithms to analyse facial expressions in real-time, with applications in for example marketing, automotive, and gaming. In marketing, companies in the US, Europe, and Asia

use Affectiva's Emotion AI to analyse viewers' reactions to advertisements, films, and TV programs. The system evaluates viewers' facial expressions and emotions to optimise content and the effectiveness of advertising campaigns. Another program is SPOT (Screening of Passengers by Observation Techniques), which detects threats at airports by analysing passengers' behavior and microexpressions. Other AI recognition programs can be found in recruitment, by companies such as Intel, which during job interviews try to detect a certain emotion based on facial expressions, then they link the assumed emotion with temperament characteristics, and later with the competencies that the company is looking for in an employee.

But why has the idea that there is a small set of universal emotions, easily interpreted from facial expressions, become so widely accepted in the field of artificial intelligence, despite numerous pieces of evidence to the contrary? It is said that it offered a small set of principles that could be applied everywhere, a simplification of complexity that was easily replicable and it can help in influencing behaviour in training people to perform in recognisable ways.

In my presentation, I begin with current contemporary applications based on Ekman's work. Later, I show that behind this there are specific assumptions that can be called emotional essentialism, which assumes that different emotions have their universal, ideal essences, understandable to every human being. Next, I move on to critical aspects of essentialist theories, with particular emphasis on Ekman's theory and, consequently, industries that use AI. I will focus on the criticism of researchers such as social scientists Maria Gendron and Lisa Feldman Barrett, and AI expert Kate Crawford. I conclude my paper with the most important question and possible answers: Can any value be found in Ekman's concept, despite the criticism? Ekman has revised his original claims over time. In this paper, I address the motivations underlying these revisions and assess the extent to which they justify the claim that his theory is consistent with experimental data. I argue that this is not the case, which supports competing paradigms in emotion theory, in particular the constructivist approach developed, among others, by Lisa Feldman Barrett.

References:

1. Crawford, K. (2021). Chapter V. *Atlas of AI. Power, Politics and the Planetary Costs of Artificial Intelligence (pp. 151-180),*
2. Coppini, S., Lucifora, C., Vicario, C. M., & Gangemi, A. (2023). *Experiments on real-life emotions challenge Ekman's model. Scientific Reports, 13*, Article 36201.
3. Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). *Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest, 20*(1), 1–68.