UABDIVULGA
BARCELONA RECERCA I INNOVACIÓ

27/01/2021

Anthropomorphic Transfer to Artificial Intelligence

Can a robot have emotions or understand those of humans? In this article Jordi Vallverdú, lecturer at the Department of Philosophy, explains with nuances a narrative that is very conducive to science fiction cinema, but which is also increasingly taking the form of an "urban legend" that is enriched by the human need to perceive reality emotionally, and come to believe that machines can feel, the sky can become angry and that turtles understand us.

"Emotional Artificial Intelligence," "Emotional Robots," or "affective computing" are terms we listen to more and more often in the various media. They all refer to the same reality: the search that unites emotions and machines, projecting a scenario that worries many people, that of machines that fully understand human emotions or, even more, have emotions for themselves. Now, what is true about all this? Are there really machines that have emotions? Despite what we have seen, the answer is clear and direct: no. However, numerous studies are advertised teaching AI or robotics systems that purport to experience emotions. This misappropriation of capabilities is at the same time the result of an anthropomorphic transfer. On the other hand, we are witnessing a second major problem in the design of machines that capture human emotions or, presumably, reproduce them: starting from a biased or directly erroneous model of what is a set of affective processes surrounding human culture.

To begin with, these fields of research are completely dominated by Western men who project their own and personal views of affective environments, so many design errors can be identified in relation to gender issues. Such a lack of interdisciplinary work leads to the use of simplistic models of the reality of emotions, such as the use of Ekman's model of 6

basic emotions, which is neither correct nor universally distributed ... but correctly fits expectations. programming and work of an engineer.

Apart from this, as our article shows, there is also an overestimation of the benefits of implementing emotional machines, something that is not entirely clear, let alone in the current state of (in) maturity of AI systems. And, paradoxically, the opposite situation also occurs with similar harmful effects: the underestimation of the power that machines with emotional abilities have for human society. We know that the designs of kind but assertive artificial voices get users to follow their directions, more so than in the case of voices with signs of insecurity.

Finally, another problem is also easily identifiable: the incorrect attribution of emotions to machines that perform mundane operations. We think of a robot that frowns and closes its mouth ... Can we say that it is angry? Or rather it seems to us? Given the intrinsic need of humans to perceive reality emotionally, we need to believe that machines feel, that the sky gets angry when it rains, or that our turtle is a pet that understands us. In conclusion, it must be admitted that in the design of machines and software that has emotional characteristics, numerous errors, biases and unfounded assumptions occur. Given that emotions make us exactly what we are, as individuals and as a species, it is vitally important to control how we introduce them into the machines that accompany us. And now, if we have aroused your interest, you will read the article. Pure emotion.

Jordi Vallverdú
Area of Logic and Philosophy of Science.
Philosophy Department.
Universitat Autònoma de Barcelona.
jordi.vallverdu@uab.cat

References

Jordi Vallverdù, Valentina Franzoni and Alfredo Milani. 2019. **Errors, Biases and Overconfidence in Artificial Emotional Modeling**. In *Proceedings of WI '19: IEEE/WIC/ACM International Conference on WebIntelligence (WI '19 Companion)*, New York, NY, USA, 86-90 pages. https://doi.org/10.1145/3358695.3361749

View low-bandwidth version