CURMUDGEON CORNER

What is a Turing test for emotional AI?

Manh-Tung Ho[1,2,3]

Emotional artificial intelligence (AI) is a subset of AI systems specializing in recognizing or even responding to human emotions. Companies all over the world are now developing these systems based on various forms of data: text, voice tone, biometrics, etc. Emotional AI, in its present form, is a weak, and narrow form of AI since it is limited by its pre-programs and it does not have the capacity to understand or experience any parts of its information processing, whether input, output, or its algorithm.

It is arguable that at some point in the future, emotional AI will achieve the status of being strong and general, which means, it is no longer limited by its programs, and it can have subjective experience of the emotions it is trained to recognize. In other words, it will pass the Turing test for emotional intelligence. In this essay, I argue for *five specifications* of the Turing test for emotional AI, drawing from Schwaninger (2022)'s work on a philosophizing machine.

Schwaninger (2022) develops a specification of the Turing test based on his observations of large language models. Here, the author specifies that the Turing test for large-language-model AI is whether it can philosophize, provided three requirements. First, there is a need to control its training data, i.e., knowing a reasonable level of detail what the training data contain and how the machine might manipulate symbols/texts to come up with its answer. Second, testing the machine to see if it has any gasp of vagueness such as in the sorites paradox. Third, the test must also cover whether the machine can come up with a psychological question, i.e., a question that identifies why humans are inclined to accept the truth of an obviously false conclusion given its induction steps and the premises.

What are the specifications of a Turing test for emotional AI then? Drawing on Schwaninger (2022)'s work, one can extrapolate the specifications of emotional AI's Turing test in a few interesting ways.

First, having a conversation about emotions is a good way to test emotional understanding of AI. Clearly, dialogues play an important role in the original Turing test as well as Schwaninger's specifications. In the case of emotional AI, given the recent increased reliance on multi-modalities of data (texts, voice tone, biometric data, video images, etc.) to develop emotional AI, it is likely that future emotional AI systems will be able to use conversation to convey its understanding of emotions. An example of a Turing test for emotional AI includes showing the AI and a control human subject a video of humans interacting, then letting an examiner pose questions to both the AI system and the human about the emotions that can be inferred from the videos. This leads to the second requirement.

Second, it is necessary to control the training data and the training protocol for emotional AI. Specifically, similar to Schwaninger's first requirement, the AI shall not have prior knowledge of certain emotions and it is necessary to know in reasonable details how such an AI system come up with an answer when being asked to recognize an emotion. When encountering emotions that are not in its training data, and whether it can realize that it does not know such emotions would provide evidence for its capacity of strong and general emotional intelligence.

Third, one can also leverage cultural differences in emotional expression to test its understanding. For example, while the AI only receives training emotional data from people in a culture, in the Turing test, an examiner can show the EAI video tapes or chats of people from a different culture. If the EAI system identifies confusion in itself, then we can say this can also be evidence of its emotional understanding.

Fourth, causal relationships among emotion, reason, and action (words spoken included) can also be leveraged to test emotional understanding of an AI system. An emotional AI

✉ Manh-Tung Ho
tung.homanh@phenikaa-uni.edu.vn

[1] Graduate School of Asia Pacific Studies, Ritsumeikan Asia Pacific University, Beppu, Oita, Japan

[2] Institute of Philosophy, Vietnam Academy of Social Sciences, Hanoi, Vietnam

[3] Centre for Interdisciplinary Social Research, Phenikaa University, Hanoi, Vietnam

system that passes the Turing test should be able to identify the possible causal relationship among emotions and actions of the people in a video or a dialogue that are being presented to it. For example, it should be able to make factual statements such as "person A breaks things because he/she feels angry" and also counterfactual statements such as "had this person not felt stressed, he/she would have not cursed." More importantly, it needs to be able to identify ambiguous situations, where it is not clear what is the causal direction of an emotion and an action. This is to leverage the concept of vagueness in Schwaninger (2022)'s work.

Finally, an emotional AI system that passes the Turing test should be required to have an intelligible conversation to philosophize about the nature of emotions. It must be said that it is still a heated debate whether emotions are biologically hardwired into human beings (i.e., the essentialist account) or emotions are socially constructed (i.e., the constructivist account). According to the theory of constructed emotion proposed by Lisa Feldman Barrett, emotions are abstract categories, constructed as mental representations of us and the world, to fulfill five functions: meaning-making, body-regulating, action-prescribing, communication, and social influence. If this account is correct, it implies that emotion expression and emotion inference are not mere cognitive functions, but it has clear behavioral and social mandates. To develop a capacity for understanding emotions, one must interact with the physical and social world. The cases of emotional disorders among children who lack social interactions when they were infant point to a highly probable conclusion: a disembodied algorithm cannot pass the Turing test for emotional intelligence, specified above.

In conclusion, this essay puts forth some considerations on the features of a Turing test for emotional AI, which includes the control of training emotional data, the use of cultural differences in emotions, the use of dialogue, the use of causal relationships among emotion, reason, and action, as well as the philosophizing on the nature of emotions. These tools serve as initial parameters for evaluating whether an emotional AI machine has achieved the status of general and strong emotional intelligence. It also highlights the tension in theoretical debates on what emotions are, whether they are biologically hardwired or constructed,

and speculates that a disembodied affect-sensing algorithm cannot pass the Turing test for emotional intelligence. It is clear that laden in our understanding of emotions is our presumptions of what constitutes a mind and its relationship with the world (Vuong 2022). Thus, clarifying philosophical implications, including the epistemology, ontology, and ethics, of emotional AI, requires the efforts of not only theorists and scientists, but also engineers. Unraveling the mystery of the mind-technology problem is crucial for identifying ways to live well and ethically with smart technologies that not only *feel* but also *feed* our emotions.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Data availability There are no data associated with this essay.

Declarations

Conflicts of interest The author declares no conflicts of interest.

Reference

Schwaninger AC (2022) The philosophising machine—a specification of the turing test. Philosophia. https://doi.org/10.1007/s11406-022-00480-5

Vuong QH (2022) Mindsponge Theory. https://books.google.com/books?id=OSiGEAAAQBAJ