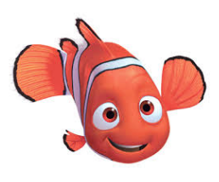
Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis

CL Journal, June 2022

PREFACE

Automatic Emotion Recognition (AER) can be a force that helps unlock:



- how emotions work; how they relate to our health, language, social interactions
- numerous commercial applications

Yet, AER can also be a tool for substantial harm:

- mass application on vulnerable populations
- unreliable approaches; privacy concerns; physiognomy

OpenGlobalRights

Strategies Topics Regions Up Close Tools Multimedia Partnerships

How emotion recognition software strengthens dictatorships and threatens democracies

Given that the idea of using emotion recognition technology as a tool of governance is an entirely flawed premise, a ban makes the most sense.

By: James Jennion Español

Should we be building AER systems? Are they ethical?

This sheet helps in thinking about these questions. It:

- documents and organizes ethical considerations
- discusses factors at play in particular contexts

Saif M. Mohammad
National Research Council Canada
⌂ http://saifmohammad.com
✉ saif.mohammad@nrc-cnrc.gc.ca @SaifMMohammad

creativity emotions fairness in language

No One Sheet to Rule them All

A single ethics sheet does not speak for the whole community

Multiple ethics sheets (by different teams, approaches) for the same or overlapping tasks can reflect multiple perspectives, viewpoints, and what is important to different groups of people at different times.

This sheet for AER is an example of “Ethics Sheets for AI Tasks” (ACL 2022)

A Call to Document Ethics Considerations at the Level of AI *Tasks*

INTRODUCTION

Scope: AER from text (AER in NLP)

Task: AER is an umbrella term for numerous tasks; e.g., inferring…

1. emotions felt by the speaker
2. emotions perceived by the listener
3. patterns of emotions over time
4. speaker’s stance to a target
5. and many more…

Tasks & Modalities come with benefits, harms, ethical considerations

Modalities for AER

- facial expressions, gait, proprioceptive data (movement of body), gestures
- skin and blood conductance, blood flow, respiration, infrared emanations
- force of touch, haptic data
- speech, text

50 ETHICAL CONSIDERATIONS

I. TASK DESIGN

A. Theoretical Foundations

1. Emotion Task and Framing
2. Emotion Models and Choice of Emotions
3. Meaning, Extra-Linguistic Information
4. Wellness and Health Implications
5. Aggregate vs. Individual Level

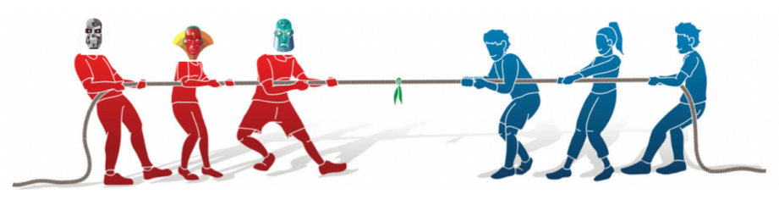
B. Implications of Automation

6. Why Automate
7. Embracing Diversity
8. Participatory Design
9. Applications, Dual Use
10. Disclosure of Automation

II. DATA

C. Why This Data

11. Types of data
12. Dimensions of data



D. Human Variability v Machine Normativeness

13. Variability of Expression, Representation
14. Norms of Emotions Expression
15. Norms of Attitudes
16. "Right" Label or Many Appropriate Ones
17. Label Aggregation
18. Historical Data
19. Training-Deployment Differences

E. The People Behind the Data

20. Platform Terms of Service
21. Anonymization and Deletion
22. Warnings and Recourse
23. Crowdsourcing

III. METHOD

F. Why This Method

24. Types of Methods and Tradeoffs
25. Who is Left Out by this Method
26. Spurious Correlations
27. Context is Everything
28. Individual Emotion Dynamics
29. Historical Behavior
30. Emotion Management, Manipulation
31. Green AI

IV. IMPACT AND EVALUATION

G. Metrics

32. Reliability/Accuracy
33. Demographic Biases
34. Sensitive Applications
35. Testing (Diverse Datasets, Metrics)

H. Beyond Metrics

36. Interpretability, Explainability
37. Visualization
38. Safeguards and Guard Rails
39. Harms when System Works as Designed
40. Contestability and Recourse
41. Be wary of Ethics Washing

V. PRIVACY, SOCIAL GROUPS

I. Implications for Privacy

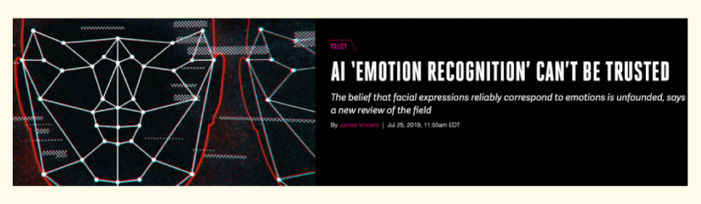
42. Privacy and Personal Control
43. Group Privacy and Soft Biometrics
44. Mass Surveillance vs. Right to Privacy, Expression, Protest
45. Right Against Self-Incrimination
46. Right to Non-Discrimination

J. Implications for Social Groups

47. Disaggregation
48. Intersectionality
49. Reification and Essentialization
50. Attributing People to Social Groups

What are the ethical considerations for your task?

1. Emotion Task and Framing



POLICY

AI 'EMOTION RECOGNITION' CAN'T BE TRUSTED

The belief that facial expressions reliably correspond to emotions is unfounded, says a new review of the field

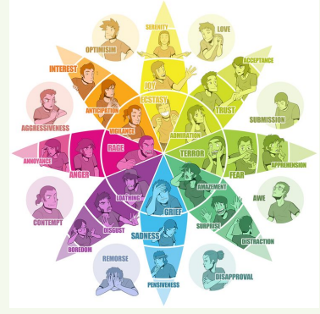
By James Vincent | Jul 25, 2019, 11:55am EDT

Is the goal to infer one’s emotions from an utterance?

- is it possible to do so?
- is it ethical to try to infer such a personal mental state?

Often, other framings are more appropriate.

2. Emotion Model and Choice of Emotions



Avoid careless endorsement of discredited ideas:

- universality of some emotions; basic emotions
- universal mapping to facial expressions (Barrett 2017)
- internal state related to outward appearance: physiognomy

8. Participatory/Emancipatory Design

“nothing about us without us”

- disabilities research (Stone and Priestley 1996)
- indigenous communities research (Hall 2014)

Center people, especially disadvantaged communities (Oliver 1997; Spinuzzi 2005, Noel 2016)

- agency to shape the design process

Paper

Poster

13-19. Human Variability v Machine Normativeness

variability in mental representation, expression of emotions
vs.
inherent bias of modern machine learning approaches
to focus on what is common (in the training data)

Through their behaviour (e.g., recognizing some forms of expressions and not others), AI systems convey to the user what is “normal”; implicitly invalidating other forms.

43. Group Privacy

Soft-biometrics

- identifying groups of people with similar traits
- people disfavour such profiling (McStay, 2020)

> There are very few Moby-Dicks. Most of us are sardines. The individual sardine may believe that the encircling net is trying to catch it. It is not. It is trying to catch the whole shoal. It is therefore the shoal that needs to be protected, if the sardine is to be saved. — Floridi (2014)