Ethics Sheet for
Automatic Emotion Recognition and Sentiment Analysis

To Appear in CL Journal, June 2022

Saif M. Mohammad
Senior Research Scientist, National Research Council Canada
Saif.Mohammad@nrc-cnrc.gc.ca @SaifMMohammad

National Research Council Canada Conseil national de recherches Canada
Canada

Automatic Emotion Recognition (AER)

A force that helps unlock:

- how emotions work
- how they relate to our health, language, behavior, social interactions,…
- numerous commercial applications that benefit society

A tool for substantial harm, e.g.:

- mass application on vulnerable populations
- unreliable approaches
- privacy concerns
- perpetuation of physiognomy

AI Tasks

- Face recognition
- Automatic Emotion Recognition (AER)
- Personality trait identification
- Machine translation
- Coreference resolution
- Image generation
- Text generation
- Text summarization
- Detecting trustworthiness
- Deception detection
- Information retrieval
- …

Female historians and male nurses do not exist, Google Translate tells its European users

by Nicolas Kayser-Bril

That Personality Test May Be Discriminating People... and Making Your Company Dumber

Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems

Svetlana Kiritchenko, Saif Mohammad

'Dangerous' AI offers to write fake news

By Jane Wakefield
Technology reporter

27 August 2019 | Comments

Should we create moral machines?

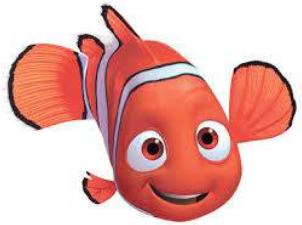
September 13, 2021 artificial intelligence, Center of Medical Ethics and Health Policy, Decision-making, Machine Ethics

What part do we play in this as researchers, system builders, leaders of tech companies?

What are the hidden assumptions in our research/work/product?

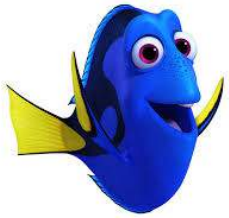
What are the unsaid implications of our choices?

my own work…



The Search for Emotions in Language

computational affective science



creativity, diversity

morality, fairness

Ethics Sheet for
Automatic Emotion Recognition and Sentiment Analysis

Medium Blog Post:
https://medium.com/@nlpscholar/ethics-sheet-aer-b8d671286682

- useful to me, and hopeful others as well
- allows stakeholders to engage

What about other AI tasks?

To Appear in CL Journal June 2022

Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis

Saif M. Mohammad*

The importance and pervasiveness of emotions in our lives makes affective computing a tremendously important and vibrant line of work. Systems for automatic emotion recognition (AER) and sentiment analysis can be facilitators of enormous progress (e.g., in improving public health and commerce) but also enablers of great harm (e.g., for suppressing dissidents and manipulating voters). Thus, it is imperative that the affective computing community actively engage with the ethical ramifications of their creations. In this paper, I have synthesized and organized information from AI Ethics and Emotion Recognition literature to present fifty ethical considerations relevant to AER. Notably, the sheet fleshes out assumptions hidden in how AER is commonly framed, and in the choices often made regarding the data, method, and evaluation. Special attention is paid to the implications of AER on privacy and social groups. Along the way, key recommendations are made for responsible AER. The objective of the sheet is to facilitate and encourage more thoughtfulness on why to automate, how to automate, and how to judge success ***well before*** *the building of AER systems. Additionally, the sheet acts as a useful introductory document on emotion recognition (complementing survey articles).*

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

No One Sheet to Rule them All

A single ethics sheet does not speak for the whole community

Multiple ethics sheets

- by different teams and approaches
- reflect multiple perspectives, viewpoints

The sheets record what is considered important to different groups of people at different times.

Ethics Sheet for
Automatic Emotion Recognition and Sentiment Analysis

...admittedly a somewhat longish document, so I will summarize

preface — introduction — applications — ethical considerations

PREFACE

Importance and complexity of emotions

Should we be building AI systems for Emotion Recognition? Is it ethical?

This sheet will…

- help in thinking about this question
- discuss factors that come into play in particular contexts
- what is more appropriate for a given context
- broaden perspective on how to assess success, etc.

INTRODUCTION

Modalities

- Facial expressions, gait, proprioceptive data (movement of body), gestures
- Skin and blood conductance, blood flow, respiration, infrared emanations
- Force of touch, haptic data (from sensors of force)
- Speech, language (esp. written text, emoticons, emojis)

All of these come with…
Benefits, Potential Harms, Ethical Considerations

Scope of this Sheet

AER from written text and AER in Natural Language Processing (NLP)

Several of the listed considerations apply to AER in general (regardless of modality, and regardless of field such as NLP or Computer Vision)

INTRO(continued): Task

1. Inferring emotions felt by the speaker
2. Inferring emotions of the speaker as perceived by the reader/listener
3. Inferring emotions that the speaker is attempting to convey
4. Inferring emotions evoked in the reader/listener
5. Inferring emotions of people mentioned in the text
6. Inferring whether what is described is good for pre-determined target of interest
7. Inferring the intensity of the emotions discussed above
8. Inferring patterns of speaker's emotions over long periods of time, across many utterances; including the inference of moods, emotion dynamics, and emotional arcs
9. Inferring speaker's emotions/attitudes/sentiment towards a target product, movie, person, idea, policy, entity, etc.
10. Inferring emotionality of language used in text (regardless of whose emotions)
11. Inferring how language is used to convey emotions such as joy, sadness, loneliness, hate, etc.
12. …

All of these come with…
Benefits, Potential Harms, Ethical Considerations

APPLICATIONS

- Public Health
  - loneliness (Guntuku et al. 2019; Kiritchenko et al. 2020)
  - depression (De Choudhury et al. 2013; Resnik et al. 2015)
  - suicidality prediction (MacAvaney et al. 2021)
  - bipolar disorder (Karam et al. 2014)
  - stress (Eichstaedt et al. 2015)
  - well-being (Schwartz et al. 2013)
- Commerce/Business
  - track sentiment towards one's products
  - develop virtual assistants, writing assistants
- Government Policy and Public Health Policy
  - tracking public opinion (e.g., towards public health measures and climate change)

Applications (continued)

- Art and Literature
  - improve our understanding of what makes a compelling story
  - what makes well-rounded characters
  - why does art evoke emotions? how do the lyrics impact us emotionally, etc.
  - can machines generate art (generate paintings, stories, music, etc.)? (Born et al. 2021)

- Social Sciences, Neuroscience, Psychology
  - what makes people thrive? What makes us happy?
  - what can our language tell us about our well-being?
  - what can language tell us about how we construct emotions in our minds?
  - what drives emotion (dys)regulation?

- Military, Policing, and Intelligence (controversial)
  - tracking misinformation

All of these come with…
Potential Harms, Ethical Considerations

Ethical Considerations

50 considerations grouped under:

- *Task Design*
- *Data*
- *Method*
- *Impact and Evaluation*

} common phases in system development

- *Implications for Privacy and Social Groups*

Task Design

Summary: This section discusses various ethical considerations associated with the choices involved in the framing of the emotion task and the implications of automating the chosen task. Some important considerations include: Whether it is even possible to determine one's internal mental state? And, whether it is ethical to determine such a private state? Who is often left out in the design of existing AER systems? I discuss how it is important to consider which formulation of emotions is appropriate for a specific task/project; while avoiding careless endorsement of theories that suggest a mapping of external appearances to inner mental states.

A. THEORETICAL FOUNDATIONS

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

B. IMPLICATIONS OF AUTOMATION

6. Why Automate (Who Benefits, Shifting Power)
7. Embracing Neurodiversity
8. Participatory/Emancipatory Design
9. Applications, Dual use, Misuse
10. Disclosure of Automation

1. Emotion Task and Framing

Carefully consider what emotion task should be the focus of the work

- make corresponding design choices

Not uncommon: One emotion task in mind and assuming that an off-the-shelf AER system is designed for that task (see variety of tasks discussed earlier)

Each task entails ethical considerations

- e.g. is the goal to infer one's emotions from an utterance?
- is it possible to do so? Is it ethical to determine such a private state?
- often, other framings are more appropriate

2. Emotion Model, Choice of Emotions

Basic Emotions Theory (BET), Dimensional Theory, Theory of Constructed Emotion,...

Discredited ideas: Universality of some emotions and their mapping to facial expressions (Barrett 2017a; Barrett et al. 2019)

Avoid careless endorsement of discredited theories, perpetuation of harmful ideas

(such as suggesting one can determine internal state from outward appearance—physiognomy)

Data

Summary: This section has three broad themes: implications of using datasets of different kinds, the tension between human variability and machine normativeness, and the ethical considerations regarding the people who have produced the data. Notably, I discuss how on the one hand there is tremendous variability in human mental representation and expression of emotions, and on the other hand, is the inherent bias of modern machine learning approaches to ignore variability. Thus, through their behaviour (e.g., by recognizing some forms of emotion expression and not recognizing others), AI systems convey to the user what is "normal"; implicitly invalidating other forms of emotion expression.

C. WHY THIS DATA

11. Types of data
12. Dimensions of data

D. HUMAN VARIABILITY VS. MACHINE NORMATIVENESS

13. Variability of Expression and Mental Representation
14. Norms of Emotions Expression
15. Norms of Attitudes
16. One "Right" Label or Many Appropriate Labels
17. Label Aggregation
18. Historical Data (Who is Missing and What are the Biases)
19. Training-Deployment Differences

E. THE PEOPLE BEHIND THE DATA

20. Platform Terms of Service
21. Anonymization and Ability to Delete One's information
22. Warnings and Recourse
23. Crowdsourcing

13-19. Point of Tension

variability in human mental representation and 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 emotion expression and not others), AI systems convey to the user what is "normal"; implicitly invalidating other forms of emotion expression.

Method

Summary: This section discusses the ethical implications of doing AER using a given method. It presents the types of methods and their tradeoffs, as well as, considerations of who is left out, spurious correlations, and the role of context. Special attention is paid to green AI and the fine line between emotion management and manipulation.

F. WHY THIS METHOD

24. Types of Methods and their Tradeoffs
25. Who is Left Out by this Method
26. Spurious Correlations
27. Context is Everything
28. Individual Emotion Dynamics
29. Historical Behavior is not always indicative of Future Behavior
30. Emotion Management, Manipulation
31. Green AI

Impact and Evaluation

Summary: This section discusses various ethical considerations associated with the evaluation of AER systems (The Metrics) as well as the importance of examining systems through a number of other criteria (Beyond Metrics). Notably, this latter subsection discusses interpretability, visualizations, building safeguards, and contestability, because even when systems work as designed, there will be some negative consequences. Recognizing and planning for such outcomes is part of responsible development.

G. METRICS

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

H. BEYOND METRICS

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

Implications for Privacy and for Social Groups

Summary: This section presents ethical implications of AER for privacy and for social groups. These issues cut across Task Design, Data, Method, and Impact. The privacy section discusses both individual and group privacy. The idea of group privacy becomes especially important in the context of soft-biometrics determined through AER that are not intended to be able to identify individuals, but rather identify groups of people with similar characteristics. The subsection on social groups discusses the need for work that does not treat people as a homogeneous group (ignoring group differences and implicitly favoring the majority group) but rather values disaggregation and explores intersectionality, while minimizing reification and essentialization of social constructs such as race and gender.

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

43. Group Privacy

Soft-biometrics and emotions

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)

People disfavour such profiling (McStay, 2020)

Circling back to Design

8. Participatory/Emancipatory Design

"nothing about us without us"

- disabilities research (Stone and Priestley 1996; Seale et al. 2015)
- indigenous communities research (Hall 2014)
- autism spectrum research (Fletcher-Watson et al. 2019)
- neurodiversity research (Brosnan et al. 2017)
- …

8. Participatory/Emancipatory Design

Centers people, especially marginalized and disadvantaged communities

- not passive subjects
- have agency to shape the design process

Spinuzzi 2005, Humphries, Mertens, and Truman 2020; Noel 2016; Oliver 1997

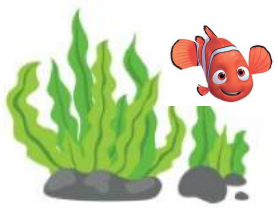
AI/NLP in service of other communities (AI4X, NLP4X)

- Led by psychologists, linguists, social scientists, clinicians,…
- AI for health, NLP for psychology, NLP for Affective Science, NLP for humanities…

What are the ethical considerations for your task?

Papers, Slides, Poster, Emotion Resources
Available at: www.saifmohammad.com

Saif M. Mohammad
Saif.Mohammad@nrc-cnrc.gc.ca
@SaifMMohammad



NRC·CNRC

Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis

CL Journal, June 2022

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

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

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

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

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)

National Research Council Canada | Conseil national de recherches Canada
Canada