Summary Card of Ethics Sheet **Task: Automatic Emotion Recognition (AER)**

Created by: Saif M. Mohammad (with input from various others) Contact: saif.mohammad@nrc-cnrc.gc.ca
Date of publication: July 2021, Date last updated: July 2021

Full sheet available at: https://medium.com/@nlpscholar/ethics-sheet-aer-b8d671286682

Primary Motivation

To create a go-to point for a carefully compiled critical engagement with the ethical issues relevant to emotion recognition; going beyond individual systems and drawing on knowledge from a body of past work.

Process

This sheet began as a way to organize my thoughts around responsible emotion recognition research based on literature review and discussions with others. Earlier drafts were sent to scholars from computer science, psychology, linguistics, neuroscience, social science, etc. Their comments helped shape the sheet.

Target Audience

The primary audience for this sheet are researchers, developers, and educators from NLP, ML, AI, data science, public health, psychology, etc. that build, make use of, or teach about AER technologies; however, much of the discussion is accessible to all stakeholders of AER.

Scope

This sheet focuses on AER from written text (in Natural Language Processing). Many considerations apply broadly to various modalities. Several considerations apply to AER (regardless of modality).

Sections

Preface: frames the discussion and presents key information about the sheet

Modalities & Scope: lists common modalities of AER data; sets the scop.

Task: lists common AER task framings and introduces how they have ethical implications

Applications: lists example applications of AER in public health, commerce, research, art, etc.

Ethical Considerations: Presents 50 ethical considerations grouped by associated development stage:

- Task Design: Theoretical foundations (5), Implications of automation (5)
- Data: why this data (2), human variability vs. machine normativeness (9), people behind data (4)
- Method: why this method (8)
- Impact and Evaluation: Metrics (4), Beyond Metrics(6)
- Privacy & Social Groups: Implications for privacy (5), Implications for Social Groups (4)

List of Ethical Considerations

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 Design and Framing
2. What Aspect of the Emotional Experience
3. Meaning and Extra-Linguistic Information
4. Wellness and Emotion
5. Aggregate Level vs. Individual Level

B. IMPLICATIONS OF AUTOMATION

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

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

1. Types of data
2. Dimensions of data

D. HUMAN VARIABILITY VS. MACHINE NORMATIVENESS

3. Variability of Expression and Mental Representation
4. Norms of Emotions Expression
5. Norms of Attitudes
6. One "Right" Label or Many Appropriate Labels
7. Label Aggregation
8. Historical Data (Who is Missing and What are the Biases)
9. Training-Deployment Differences

E. THE PEOPLE BEHIND THE DATA

10. Platform Terms of Service
11. Anonymization and Ability to Delete One's information
12. Warnings and Recourse
13. Crowdsourcing

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. I also discuss green AI and the fine line between emotion management and manipulation.

F. WHY THIS METHOD

1. Types of Methods and their Tradeoffs
2. Who is Left Out by this Method
3. Spurious Correlations
4. Context is Everything
5. Individual Emotion Dynamics
6. Historical Behavior is not always indicative of Future Behavior
7. Emotion Management, Manipulation
8. 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

1. Reliability/Accuracy
2. Demographic Biases
3. Sensitive Applications
4. Testing (on Diverse Datasets, on Diverse Metrics)

H. BEYOND METRICS

5. Interpretability, Explainability
6. Visualization
7. Safeguards and Guard Rails
8. Harms even when the System Works as Designed
9. Contestability and Recourse
10. 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

1. Privacy and Personal Control
2. Group Privacy and Soft Biometrics
3. Mass Surveillance vs. Right to Privacy, Expression, Protest
4. Right Against Self-Incrimination
5. Right to Non-Discrimination

J. IMPLICATIONS FOR SOCIAL GROUPS

6. Disaggregation
7. Intersectionality
8. Reification and Essentialization
9. Attributing People to Social Groups