Asian Bioethics Review (2023) 15:417–430
https://doi.org/10.1007/s41649-022-00237-y

ORIGINAL PAPER

The Ethics of Emotional Artificial Intelligence: A Mixed Method Analysis

Nader Ghotbi[1]

Received: 14 October 2022 / Revised: 19 November 2022 / Accepted: 22 November 2022 /
Published online: 2 December 2022

Abstract

Emotions play a significant role in human relations, decision-making, and the motivation to act on those decisions. There are ongoing attempts to use artificial intelligence (AI) to read human emotions, and to predict human behavior or actions that may follow those emotions. However, a person's emotions cannot be easily identified, measured, and evaluated by others, including automated machines and algorithms run by AI. The ethics of emotional AI is under research and this study has examined the emotional variables as well as the perception of emotional AI in two large random groups of college students in an international university in Japan, with a heavy representation of Japanese, Indonesian, Korean, Chinese, Thai, Vietnamese, and other Asian nationalities. Surveys with multiple close-ended questions and an open-ended essay question regarding emotional AI were administered for quantitative and qualitative analysis, respectively. The results demonstrate how ethically questionable results may be obtained through affective computing and by searching for correlations in a variety of factors in collected data to classify individuals into certain categories and thus aggravate bias and discrimination. Nevertheless, the qualitative study of students' essays shows a rather optimistic view over the use of emotional AI, which helps underscore the need to increase awareness about the ethical pitfalls of AI technologies in the complex field of human emotions.

Keywords Affective computing · Artificial intelligence · AI bias · AI discrimination · AI ethics · Emotional AI

✉ Nader Ghotbi
nader@apu.ac.jp

[1] College and Graduate School of Asia Pacific Studies, Ritsumeikan Asia Pacific University, Beppu City, Japan

Introduction

Emotions used to be considered a hindrance to rational decision-making and reasoning; however, research has demonstrated the important role they play in proper decision-making and satisfaction from life (Lerner et al. 2015). Therefore, emotions became the target of new research by academics from many disciplines who have studied, defined, and classified various emotions and examined the role of cognitive and social processes which shape universal and culture-specific emotions, respectively (Keltner and Lerner 2010). Burton (2015) differentiated between an emotional experience, brief and episodic, an emotion, continuing for a long time, and a trait which is a disposition to have certain emotions. Such differentiation is necessary as it has significant implications for emotional recognition by AI. For example, a job interview to find an applicant with a pleasing (grateful) character may end up picking the wrong person who appeared more pleasing at a few moments during the interview. It is probably immature to expect that emotional expressions during an interview are a reliable predictor of character in the long run. Correspondence bias refers to making such mistakes in judgement, as Scopelliti et al. (2018) demonstrated how "people infer stable personality characteristics from others' behavior, even when that behavior is caused by situational factors."

Nowadays, people express their emotions in so many ways, from facial expression, changes in voice and body gesture to verbal comments on social media, and the use of emoticons and memes (Terzimehić et al. 2021). AI applications developed by business companies have used cameras to capture facial expressions for the recognition and categorization of emotions (McStay 2020). However, Barrett et al. (2019) have criticized the common view that facial expressions of emotions are universal and can reliably be used to infer specific emotions by other people. Fernández-Dols and Russel (2017) reviewed the current psychology of facial expression, and Ekman (2017) reviewed studies that had examined the universality of the expression of emotions among various cultures around the world; he concluded that the evidence supports the universality of the expression of "happiness, anger, disgust, sadness, and fear/surprise". However, he acknowledged that cultural display rules of emotions may be different, and people may inhibit or fabricate the expression of emotions (Ekman 2017).

AI algorithms can also examine emotions by focusing on the used language. The choice of words in a written text, the symbols used in online comments and messages, or transcripts of audio conversations in speech may be used to recognize emotions (Pang and Lee 2008; Strapparava and Mihalcea 2008). Some recruitment companies have used the recording of job interviews for analysis by AI to choose the right candidates for various positions (Zetlin 2018). Greene (2020) recommends research to explore the ethical issues over unintended consequences of creating and deploying emotional AI technologies to sense, recognize, influence, and simulate human emotions and affect. He refers to a vast range of benefits in the use of emotional AI as in the detection and treatment of illnesses, assistance in disabilities, social robots for home care, chatbots for mental

healthcare, automotive and industrial safety, education, animal farming, and law enforcement and detection of threats while recognizing the risks and possible ethical harms to privacy and other civil rights, human autonomy, transparency, accuracy and inclusivity, and a lack of legal frameworks to catch up with the fast pace of emotional AI development (Greene 2020).

This study was planned to examine the potential ethical pitfalls of emotional AI by collecting emotional data from large groups of 18- to 24-year-old college students who consented to participate in the study. It thus first examined a random group of 124 college students in an international university in Japan regarding some of the emotions they felt, expressed to others, or suppressed, and searched for meaningful patterns in the collected data and whether there were reliable associations to help predict those patterns. Next, another larger group of 235 college students from the same university was invited to write an essay about the ethical use of emotional AI, and the content of their essays was analyzed with qualitative methods to understand their attitudes and reasons to support or refute the use of emotional AI applications in light of their ethical risks of harm versus potential benefits. The results of both the quantitative survey and the qualitative analysis are discussed.

Research Methods

The study included a survey questionnaire about emotions felt, suppressed, or expressed by the respondents in an ordinal scale, with the collected data quantitatively analyzed, as well as an essay contest, with the content qualitatively analyzed using a coding method for key concepts and phrases. The survey questionnaire inquired about the respondents' choice of words associated with 9 emotions, the frequency or intensity of feeling, suppressing, and expressing those emotions on an ordinal 5-level Likert scale, gender, age, nationality, and the level of religiosity (see Table 1). The responses were collected anonymously on a Google Document form and then examined using Excel and the SPSS software 27 edition, for descriptive

Table 1 The questions in the questionnaire study

A-1	Write 3 to 5 words that you associate with the following emotions: Joy/Happiness, Anger, Sadness, Fear, Surprise, Disgust, Shame, Love/Care, Lust
-2	How often or how intensely do you feel those emotions? (1 to 5 scale)
-3	Would you try to control/suppress those emotions when you feel them? (1 to 5 scale)
-4	How easily do you express (don't mind showing) those emotions to others? (1 to 5 scale)
B-1	In dealing with life problems in general, do you depend more on how you feel (emotional decision) or more on how you reason (rational decision) about them? (1 to 5 scale)
-2	How religious are you? (1 to 5 scale)
-3	Your biological gender? (female, male)
-4	Your nationality?
-5	Your age?

and inferential statistics, respectively. Table 1 summarizes the questions in the survey corresponding to the fields of data collected by the form.

For the essay contest, the students were asked to first read educational material about the ethics of emotional AI (from: https://partnershiponai.org/paper/the-ethics-of-ai-and-emotional-intelligence/) and then to decide whether they were more hopeful of the positive uses of emotional AI or more concerned about its negative outcomes, and to explain their arguments in about 1000 words (see Table 2). The essays were carefully examined, and students' choices and arguments were extracted and coded so that similar arguments could be grouped under a common code. The extracted codes were double-checked and were excluded if they were about AI in general, but not "emotional" AI. The final extracted codes are listed in the "Results" section.

Results

The anonymous respondents to the survey questionnaire included 124 college students between 18 to 24 years old with an average age of 20.5 years old. Almost half of the students ($n=61$, %49) were from Japan, with the rest from Indonesia ($n=14$, %11), Korea ($n=12$, %10), China ($n=11$, %9), Thailand ($n=10$, %8), Vietnam ($n=7$, %6), and a few from other countries. As for gender, 80 (%65) of the respondents were female and 44 (%35) were male.

Figure 1 depicts the 5-level Likert scales of how frequently and/or intensely the 9 emotions were felt by the respondents, from rarely/barely (scale 1) to very often/strongly (scale 5). Three patterns can be recognized in these charts; *joy* and *love* dominate the right side of the chart with the ordinal scales 4 and 5 getting the highest number of hits; *anger*, *disgust*, *surprise*, and *shame* dominate the left side of the chart with ordinal scales 1 and 2 getting the highest number of hits; *sadness*, *lust*, and *fear* dominate the middle part of the chart with ordinal scales 2, 3, and 4 getting the highest number of hits. These findings may be interpreted as *happiness* and *love* being felt more often among a group of young and relatively healthy college students, as compared with *anger*, *disgust*, *surprise*, and *shame*; however, as a

Table 2 Instructions for the essay question

1-	Spend some time reading the material about emotional AI and its pros and cons from: https://partnershiponai.org/paper/the-ethics-of-ai-and-emotional-intelligence/
2-	Which one of the following statements do you support? Pick a side:
P1:	*"If artificial intelligence can help individuals better understand and control their own emotional and affective states, there is enormous potential for good and a better quality of life."**
P2:	*"If artificial intelligence can automate the ability to read or control others' emotions, it has substantial implications for economic and political power and individuals' rights." **
	Write an essay to support your position in about 1000 words

*These sentences are excerpts from the same web page: https://partnershiponai.org/paper/the-ethics-of-ai-and-emotional-intelligence/

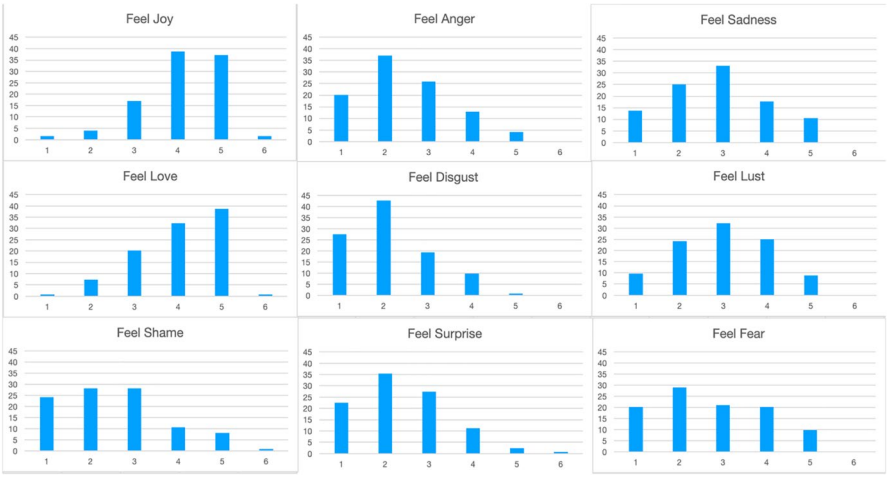


Fig. 1 Results of the 5-level Likert scales of how frequently and/or intensely the 9 emotions were felt by the respondents, from rarely/barely (scale 1) to very often/strongly (scale 5); number 6 refers to missing responses which are only a few

stressed-out group of college students dealing with many challenges of their age and study, they may be prone to feeling some *sadness* and *fear*, as well as *lust*.

However, many of the respondents would attempt to suppress their emotions to some extent (Fig. 2), with *anger*, *sadness*, *shame*, and *fear* being suppressed more often as compared with *joy*, *love*, and *surprise*. It is likely that the level of suppression of emotions depends on how negatively they are considered by the respondents under sociocultural influences, and possibly, this pattern follows

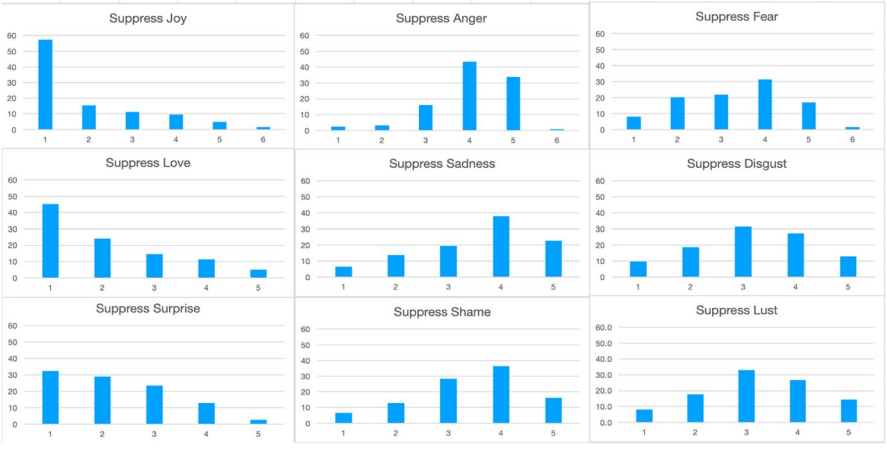


Fig. 2 Results of the 5-level Likert scales of how much the respondents would attempt to suppress the 9 emotions they felt, from not at all (scale 1) to very much (scale 5); number 6 refers to missing responses which are only a few

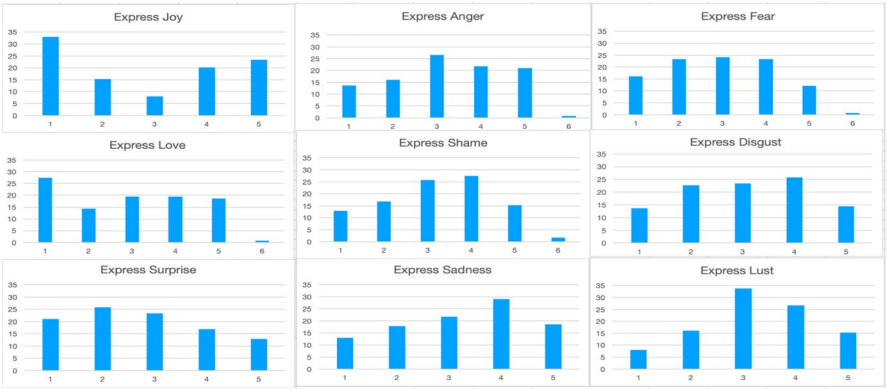


Fig. 3 Results of the 5-level Likert scales of how easily the respondents would express to others the 9 emotions they felt, from not easily (scale 1) to very easily (scale 5); number 6 refers to missing responses which are only a few

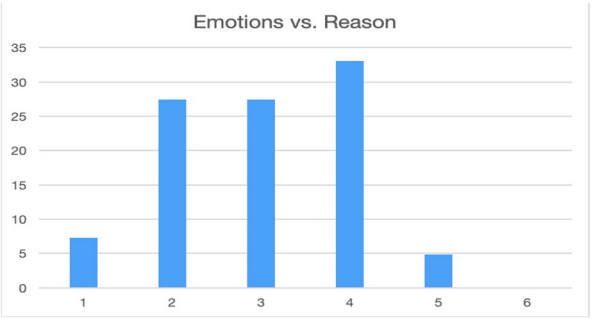


Fig. 4 Results of the 5-level Likert scales of how much decision-making depends on emotions or reasoning, from mostly on emotion (scale 1) to mostly on reason (scale 5)

the social norms whereby certain emotions are commonly described as positive and certain others are described as negative emotions. The Likert scales of *lust* and *disgust* are more pronounced in the middle to the right side of their associated chart which may be interpreted as respondents considering them either less controllable (*disgust*) or less negatively (*lust*) in comparison with *anger*, *sadness*, and *shame*.

Furthermore, when it comes to the expression of the emotions to others, there is another quite different pattern (Fig. 3). Even emotions commonly referred to as positive and less suppressed by the respondents, such as *joy* and *love*, may not be expressed and shown to others as often as they are felt. Most other emotions dominate close to the middle of the Likert scale which may be interpreted as respondents attempting to tone down the expression of their emotions to other people. Interestingly, it seems to be easier for the respondents to express *lust* rather than *love*, though the questionnaire did not inquire to whom the emotion would be expressed, one's friends or a romantic partner.

The responses to the question of whether decision-making depends on emotions or reasoning are shown in Fig. 4, which demonstrates that most respondents

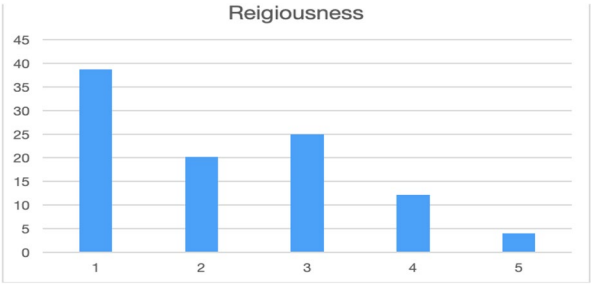


Fig. 5 Results of the 5-level Likert scales of how religious the respondents are, from not religious at all (scale 1) to very religious (scale 5)

Correlations

			V39	V14
Kendall's tau_b	V39	Correlation Coefficient	1.000	.164*
		Sig. (2-tailed)	.	.027
		N	124	124
	V14	Correlation Coefficient	.164*	1.000
		Sig. (2-tailed)	.027	.
		N	124	124
Spearman's rho	V39	Correlation Coefficient	1.000	.199*
		Sig. (2-tailed)	.	.026
		N	124	124
	V14	Correlation Coefficient	.199*	1.000
		Sig. (2-tailed)	.026	.
		N	124	124

*. Correlation is significant at the 0.05 level (2-tailed).

Feeling Fear and Religiousness

Correlations

			V39	V24
Kendall's tau_b	V39	Correlation Coefficient	1.000	.210**
		Sig. (2-tailed)	.	.005
		N	124	123
	V24	Correlation Coefficient	.210**	1.000
		Sig. (2-tailed)	.005	.
		N	123	123
Spearman's rho	V39	Correlation Coefficient	1.000	.254**
		Sig. (2-tailed)	.	.005
		N	124	123
	V24	Correlation Coefficient	.254**	1.000
		Sig. (2-tailed)	.005	.
		N	123	123

**. Correlation is significant at the 0.01 level (2-tailed).

Expressing Fear and Religiousness

Correlations

			V39	V22
Kendall's tau_b	V39	Correlation Coefficient	1.000	.206**
		Sig. (2-tailed)	.	.006
		N	124	123
	V22	Correlation Coefficient	.206**	1.000
		Sig. (2-tailed)	.006	.
		N	123	123
Spearman's rho	V39	Correlation Coefficient	1.000	.253**
		Sig. (2-tailed)	.	.005
		N	124	123
	V22	Correlation Coefficient	.253**	1.000
		Sig. (2-tailed)	.005	.
		N	123	123

**. Correlation is significant at the 0.01 level (2-tailed).

Expressing Anger & Religiousness

Correlations

			V39	V23
Kendall's tau_b	V39	Correlation Coefficient	1.000	.181*
		Sig. (2-tailed)	.	.014
		N	124	124
	V23	Correlation Coefficient	.181*	1.000
		Sig. (2-tailed)	.014	.
		N	124	124
Spearman's rho	V39	Correlation Coefficient	1.000	.215*
		Sig. (2-tailed)	.	.017
		N	124	124
	V23	Correlation Coefficient	.215*	1.000
		Sig. (2-tailed)	.017	.
		N	124	124

*. Correlation is significant at the 0.05 level (2-tailed).

Expressing Sadness & Religiousness

Correlations

			V39	V26
Kendall's tau_b	V39	Correlation Coefficient	1.000	.172*
		Sig. (2-tailed)	.	.020
		N	124	124
	V26	Correlation Coefficient	.172*	1.000
		Sig. (2-tailed)	.020	.
		N	124	124
Spearman's rho	V39	Correlation Coefficient	1.000	.206*
		Sig. (2-tailed)	.	.022
		N	124	124
	V26	Correlation Coefficient	.206*	1.000
		Sig. (2-tailed)	.022	.
		N	124	124

*. Correlation is significant at the 0.05 level (2-tailed).

Expressing Disgust & Religiousness

Correlations

			V39	V38
Kendall's tau_b	V39	Correlation Coefficient	1.000	.151*
		Sig. (2-tailed)	.	.043
		N	124	124
	V38	Correlation Coefficient	.151*	1.000
		Sig. (2-tailed)	.043	.
		N	124	124
Spearman's rho	V39	Correlation Coefficient	1.000	.177*
		Sig. (2-tailed)	.	.049
		N	124	124
	V38	Correlation Coefficient	.177*	1.000
		Sig. (2-tailed)	.049	.
		N	124	124

*. Correlation is significant at the 0.05 level (2-tailed).

Suppressing Lust & Religiousness

Fig. 6 Results of the correlation between religiousness and feeling fear, expressing fear, sadness, anger, and disgust, and suppression of lust, using the Spearman correlation coefficient (rho) and the Kendall correlation coefficient (tau)

consider both their emotions and logical reasoning when making important decisions, as most of the hits are in the middle of the Likert scale rather than the left or the right side of the Likert scale. The number of respondents who mainly consider their emotions (scale 1), or mainly reasoning (scale 5), is very small.

The respondents were also asked about their level of religiousness, and the responses show that most students in our sample were not religious (Fig. 5); the number of responses on scales 4 and 5 is the smallest though 25% of respondents picked the 3rd scale, in the middle (somewhat religious).

Although the number of very religious students in the sample was small, we looked for correlations between the level of religiousness and the expression or suppression of certain emotions such as love and lust using the Spearman correlation coefficient (rho) and the Kendall correlation coefficient (tau) on SPSS software. The analysis showed a statistically significant correlation between religiousness and feeling fear, between religiousness and expressing fear, sadness, anger, and disgust, and between religiousness and suppression of lust (Fig. 6).

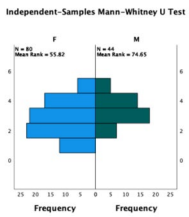
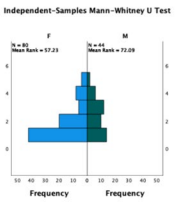
Feeling Lust and Gender	**Independent-Samples Mann-Whitney U Test Summary** Total N: 124 Mann-Whitney U: 2294.500 Wilcoxon W: 3284.500 Test Statistic: 2294.500 Standard Error: 185.168 Standardized Test Statistic: 2.887 Asymptotic Sig.(2-sided test): .004	
Suppressing Love and Gender	**Independent-Samples Mann-Whitney U Test Summary** Total N: 124 Mann-Whitney U: 2182.000 Wilcoxon W: 3172.000 Test Statistic: 2182.000 Standard Error: 180.563 Standardized Test Statistic: 2.337 Asymptotic Sig.(2-sided test): .019	

Fig. 7 Results of the correlation between gender (female vs. male) and feeling lust, and between gender and suppressing love. As seen in the graphs on the right side, the results indicate that males (M) feel more lust and suppress love more often than females (F)

Fig. 8 The word clouds for the two emotions of joy (left) and love (right). Some of the frequent words are common to both emotions, such as *family* and *friend*, and others appear as more specific to each emotion, such as *fun*, *smile*, *laugh* in Joy, and *hug*, *adore*, and *affection* in love

We examined if there were any correlations between gender and the responses to questions about emotions using the Mann–Whitney *U* test, and the analysis determined a significant correlation only for feeling lust and suppression of love, with males feeling more lust but suppressing love in comparison with females, as shown in Fig. 7.

The correlations found between nationality and the collected variables on emotions are not being reported out of ethical concerns over bias and discrimination. This issue will be explained further in the "Discussion" section. The words that had been written by the respondents corresponding to the 9 emotions were examined with the help of a text mining application to look for the most frequent words and whether there was a clear pattern to help recognize emotions in text. Figure 8 shows two of the word clouds that are used to visualize the most common words associated with the emotions of *joy* and *love* as chosen by the respondents. A comparison of the two word clouds helps identify some common words as well as more specific words related to each of these two emotions. There are many software platforms that can be programmed to mine for such words and predict

the dominant emotional context. They can be adjusted for a certain language and cultural background if a large sample of associated text can be obtained and used to train the system. RapidMiner is a powerful software that can help extract the words from irrelevant elements in a multitude of text formats, including those from social media and the Internet, stem the various grammatical forms of a word, count their frequency, and visualize a word cloud of them to help identify the context as well as the emotional tone conveyed by the text. Such identification can be done by a human observer, or an AI application that compares the choice of the words and their pattern with the corresponding patterns in its database.

Table 3 presents the most frequent words associated with the 9 emotions in our collected data. A close examination of the most frequent words suggests that it may be possible to recognize the conveyed emotions based on the patterns in the selected words by a writer and to guess the emotion being carried by those words; however, there are nuances in the process as seen in the words collected from the respondents. First, several words are common to more than one emotion, such as *family, friend*, *failure*, *cry*, and *love*. Second, the choice of words may depend on other factors such as the level of literacy (for example, several students had spelling errors that needed correction for proper classification), as well as cultural and personal variations. Third, the respondents provided their own relatively unique selection of words for each emotion, and although many words were used more commonly as a whole, there were many other words that were used by only a small number of respondents. Therefore, the accuracy level of an AI application in emotion recognition would depend on the complexity of the underlying algorithms, its access to samples of emotional material such as a written text previously sampled from that individual, and whether it can use deep learning methods to understand the emotion beyond just the choice and frequency of word usage.

The qualitative examination of 235 essays demonstrated that 137 (58%) students had a positive attitude, and 98 (42%) students had a negative attitude

Table 3 The 6 most frequent words associated with listed emotions, with their frequency of occurrence in the survey of 124 college students

Joy/happiness	*Anger*	*Sadness*	*Fear*	*Surprise*
Smile (40)	Frustrated (16)	Cry (46)	Ghost (25)	Shock (28)
Success (28)	Fight (14)	Lonely (31)	Horror (14)	Unexpected (24)
Family (24)	Cry (11)	Failure (25)	Terror (13)	Amaze (23)
Friend (18)	Violence (9)	Tears (16)	Panic (12)	Birthday (16)
Laugh (20)	Betrayal (6)	Death (14)	Failure (12)	Gift (13)
Love (10)	Stress (6)	Blue (11)	Death (11)	Present (10)
Disgust	*Shame*	*Love*	*Lust*	
Hatred (29)	Failure (26)	Family (42)	Desire (46)	
Dislike (20)	Shy (18)	Friend (25)	Sex (22)	
Dirty (14)	Mistake (15)	Affection (11)	Money (22)	
Vomit (10)	Embarrass (11)	Lover (8)	Love (14)	
Nausea (6)	Guilt (11)	Sympathy (7)	Greed (13)	
Cockroach (5)	Blush (9)	Adore (6)	Passion (12)	

towards the use of emotional AI. The arguments for a positive attitude towards emotional AI included the following:

(1) Development of software applications and products that benefited the society by offering new utilities and increasing the efficiency of existing ones, improving automated functions, better data analysis and support for decision-making systems, targeted marketing, better business planning, smarter operation of systems, etc.
(2) Improved assistance to people living alone or suffering from disabilities and assisting them with communication through emotion recognition
(3) Increasing the safety of driverless function through analysis of driver's state, improved driving assistance, better response to emergencies, etc.
(4) Assisting with criminal investigations through detection of deviations, identification of criminal behavior, online delinquent and terrorism activities, and fraud detection through emotion analysis, etc.
(5) Enhancing the quality of education through learning companion and support, interactive learning, adjusting difficulty levels, substitute teachers, simulation training, etc.
(6) Support of healthcare system through better monitoring, remote medical checkups, mental health support through recognition of human emotions and emotional responses, detection of emotional problems and needed counseling
(7) Provision of personalized entertainment, game development, product recommendations, market research, etc.
(8) Detection of employees' or customers' dissatisfaction, improved customer service, better-tuned machine consulting service, etc.
(9) Contribution to behavioral science by serving as a source of data, helping understand human emotions, and helping change the mindset of people by bringing in new values, for example through a nonjudgmental attitude of machines towards humans, etc.

The arguments for a negative attitude towards emotional AI included the following:

(1) Risk of harms associated with leakage of personal information and inability to protect privacy, exploitation of personal data for commercial purposes, machine learning bias, misleading information, misinterpretation and mistakes, the black box problem, and lack of transparency
(2) Possibility of misuse for political, economic, and marketing manipulation, use in surveillance of society to monitor citizens, data exploitation for commercial purposes, and aggravation of consumptive behavior, affecting public opinion and causing social disruption, vicious spread of misinformation, etc.
(3) Absence or inability to compensate for human interaction, lack of affection and morality, increased social isolation
(4) Causing an identity or existential crisis, fear of losing free will or rights, perpetuating stereotypes and presumptions, excluding cultural diversity and religious

affiliation, lack of empathy in psychological counseling, and causing psychological harm
(5) AI does not take responsibility for its actions and their consequences
(6) High costs of technology expanding the gap between the rich and the poor, and loss of jobs due to substitution for human contact

Discussion

The quantitative analysis of the emotional survey and the qualitative study of the essays shed a light on two major research areas on the ethics of emotional AI. One shows that even a relatively small number of emotional data can identify factors, such as gender, religiosity, and nationality that may be used to *classify* people into groups, which is a good example of how AI bias and discrimination may result from such analysis. Although correlation studies are common in social research, their results are interpreted cautiously and in general, they do not prove a hypothesis but can help generate a hypothesis that needs further research with more stringent criteria and evidence. Searching for correlations can be an interesting method to generate some hypotheses for social studies when conducted by researchers who are familiar with the limitations and shortcomings in the interpretation of the results. However, the classification of people into groups based on correlations found by AI algorithms in the hands of businesses and political entities may lead to an aggravation of social stigma and other discriminatory problems that exist in the society. Our relatively small sample of 124 college students demonstrated how a search for correlation having statistical significance may suggest associations that may be described as intriguing or interesting. The correlations found between nationality and emotional variables, especially Japanese vs. non-Japanese respondents, were so discriminatory that we found it unethical to report them at all. It may be said that searching for correlations, without first constructing a plausible hypothesis based on a large number of reported observations, is just a form of data dredging. Unfortunately, this is how AI may treat data, by searching for "meaningful" associations without first checking for the plausibility of the association, its limitations, and the possibility of random correlations in a large pool of data. There is also the danger of doing an autocorrelation study when the two variables being examined include a common component and are not independent factors.

Emotional data can be too complex to be assessed by automated algorithms and there are too many nuances to be researched before a reliable form of evaluation can be appropriately processed through an AI system. Moreover, emotion recognition technologies are far away from an accurate assessment of the complexities in the expression of human emotions. For instance, many problems could arise if the difference between emotional expressions (like a smile) versus emotional states (like happiness) has not been acknowledged. Our study revealed the large amount of variation in how nine emotions were felt by 124 college students, how they would attempt to suppress those emotions, and how they would attempt to express some emotions to others and hide others. Our demonstration of statistically significant correlations between the level of religiousness and the feeling of fear, the suppression

of lust, and the expression of fear, anger, sadness, and disgust was only the result of a search in data for any possible relationship. The data also supported an association between the male gender and the feeling of lust while suppressing love, which has been a common form of stereotyping. However, there are many nuances to such an interpretation. The examination of correlations between nationality and emotional data generated such discriminatory results that the author would not dare report out of ethical concerns. The approach of human researchers to the examination of data is based on research controls which include the sociocultural context, an understanding of the large overlap between groups, and the diversity and variation within the groups themselves. Research ethics requires researchers to not jump into conclusions before examining the limitations in their research methodology. Unfortunately, AI systems may easily bypass such controls and lead to biased results that can be discriminatory and untrue.

On the other hand, the qualitative assessment of the essays showed the dominance of a positive attitude among a larger group of 235 college students who had been provided with expert information about the pros and cons of emotional AI applications. Both the larger number of students who were optimistic about emotional AI usage (137 vs. 98) and the number of arguments for and against emotional AI (9 vs. 6) suggest that the educated young are generally more optimistic about the uses of emotional AI. This optimism in the face of biased and discriminatory results generated from the study of another group of students from the same university implicates that the ethical aspects of emotional AI have not been taken seriously by a larger percentage of students (58% vs. 42%). Understanding the reasons behind this naïve optimism among the majority of the students would require further study using interviews to inquire more about the knowledge, attitude, and practice of the students regarding emotional AI technologies.

Conversely, some researchers have suggested that AI applications may be less discriminatory than some human employers. For example, Zetlin (2018) claims that human recruiters may unconsciously bias against some applicants while a properly trained AI may be able to decide more objectively and help reduce human bias; Zetlin also reported that employers as well as many job applicants benefited from the convenience of AI administration of job interviews by saving resources and time, as candidates could choose the time they wanted to be interviewed, for example. The Japanese company Unilever has claimed that the use of AI for job interviews in fact contributed to ethnic diversity, with a "significant increase in non-white hires." The Japanese society appears to have an optimistic view of the use of AI and views it as simply another step in the automation of services; it helps with the relative lack of young workers in an aging population and reduces the need for personal interactions that may be culturally stressful and also carry the risk of infection in the ongoing infectious pandemic era. Is it possible that the convenience of using automated AI applications including emotional AI technologies is blinding many people to the ethical risks involved in their usage?

This study has limitations including the relatively smaller number of students from Asian countries other than Japan which implies the results cannot be generalized to those Asian countries. Moreover, emotion recognition was only tested on written text, but emotional AI technologies may also collect voice (audio),

facial features (video), and other biometric data (such as pulse, blood pressure, etc.). Although the qualitative part of the analysis included a fairly large sample size, the quantitative part was limited to only 124 students. Still, the analysis demonstrated the high risk of stereotyping, bias, and discrimination in even a relatively small size, but it would be better to test the findings in larger samples in the future.

Conclusion

This study first examined the emotional attributes of 124 college students and demonstrated how a search for statistically significant correlations in their responses could lead to ethically questionable stereotyping, bias, and discrimination. Meanwhile, the ability of text mining for emotional recognition in the word choices of the students for a variety of emotions was examined and it helped identify many nuances in the accuracy of such methods which are commonly employed in emotional AI applications. This is a small demonstration of how mistakes may follow the use of emotional recognition technologies.

Next, the detailed essays of 235 college students, who were instructed to read a concise source of information on the pros and cons of emotional AI, were examined using qualitative methods and a coding system helped classify their responses into 9 main ways emotional AI could be beneficial and 6 main ways they could cause ethical harm. The relatively higher proportion of students who supported the use of emotional AI for its potential benefits versus those who were against the deployment of emotional AI technologies (58% vs. 42%) confirms a relatively more optimistic view towards AI applications in general which may be a common attitude among Asian communities. It is hoped that the risks of ethical harm associated with the use of emotional AI applications will be studied more and the results help regulate their usage based on the principles of beneficence and non-maleficence.

Funding This study received support from the "Emotional AI in Cities: Cross Cultural Lessons from UK and Japan on Designing for an Ethical Life" funded by the JST-UKRI Joint Call on Artificial Intelligence and Society (2019).

Declarations

Ethical Approval The research project presented its research methodology to the ethical committee of the university with detailed explanation and received approval.

Consent to Participate All students were free to participate in either the survey or the essay contest and could withdraw from the study whenever they wanted. The study was done anonymously.

Consent for Publication The author and students who anonymously contributed to the study consent to publication of the research results considering the anonymity and respect to privacy of the participants.

Conflict of Interest The authors declare no competing interests.

References

Barrett, Lisa Feldman, Ralph Adolphs, Stacy Marsella, Alex M. Martinez, and Seth D. Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20: 1–68. https://doi.org/10.1177/1529100619832930.

Burton, Neel. 2015. *Heaven and hell: The psychology of the emotions*. Oxford: Acheron Press.

Ekman, Paul. 2017. Facial expressions. In *The science of facial expression*, ed. Jose-Miguel Fernández-Dols, and James A. Russell, 39–56. New York: Oxford University Press.

Fernández-Dols, Jose-Miguel, and James A. Russell (Eds.). 2017. *The science of facial expression*. New York: Oxford University Press.

Greene, Gretchen. 2020. The Ethics of AI and Emotional Intelligence: Data sources applications and questions for evaluating ethics risk. *Partnership on AI*, 30 July 2020. https://partnershiponai.org/paper/the-ethics-of-ai-and-emotional-intelligence/. Accessed 19 Nov 2022.

Keltner, Dacher, and Jennifer S. Lerner. 2010. Emotion. In *Handbook of social psychology*, ed. Susan T. Fiske, Daniel T. Gilbert, and Gardner Lindzey, 317–352. New York: John Wiley & Sons. https://doi.org/10.1002/9780470561119.socpsy001009..

Lerner, Jennifer S., Ye Li, Piercarlo Valdesolo, and Karim S. Kassam. 2015. Emotion and decision making. *Annual Review of Psychology* 66: 799–823. https://doi.org/10.1146/annurev-psych-010213-115043.

McStay, Andrew. 2020. Emotional AI and EdTech: Serving the public good? *Learning, Media and Technology* 45 (3): 270–283. https://doi.org/10.1080/17439884.2020.1686016.

Pang, Bo, and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1–2): 1–135. https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf. Accessed 19 Nov 2022.

Scopelliti, Irene, H. Lauren Min, Erin McCormick, Karim S. Kassam, and Carey K. Morewedge. 2018. Individual differences in correspondence bias: Measurement, consequences, and correction of biased interpersonal attributions. *Management Science* 64(4): 1879–1910. https://doi.org/10.1287/mnsc.2016.2668.

Strapparava, Carlo, and Rada Mihalcea. 2008. Learning to identify emotions in text. Proceedings of the 2008 ACM Symposium on Applied Computing, 1556–1560. https://web.eecs.umich.edu/~mihalcea/papers/strapparava.acm08.pdf. Accessed 19 Nov 2022.

Terzimehić, Nada, Svenja Yvonne Schött, Florian Bemmann, and Daniel Buschek. 2021. MEMEories: Internet memes as means for daily journaling. In *DIS '21: Designing Interactive Systems Conference 2021*, 538–548. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3461778.3462080.

Zetlin, Minda. 2018. AI is now analyzing candidates' facial expressions during video job interviews. *Inc.*, 28 February 2018. https://www.inc.com/minda-zetlin/ai-is-now-analyzing-candidates-facial-expressions-during-video-job-interviews.html. Accessed 19 Nov 2022.