Emotional Machines: Ethics and Biases of Emotion Artificial Intelligence in Businesses and Workplaces

Sumeet Jhamb
University of Alaska Anchorage

Teresa Ryan
University of Alaska Anchorage

This study discusses the emerging technology of emotion artificial intelligence and its ethical implications, particularly bias, in business. Emotion AI offers potential for many fields of business, but comes with inherent concerns over privacy, security, and ethics (Dattner et al., 2019). Past studies have revealed inherent gender and racial biases in artificial intelligences in many different aspects of business (Buolamwini, 2019; Feast, 2019; Tinsley and Ely, 2018; Wellner and Rothman, 2019). The present study considers emotion AI's current capabilities, recognizes its inherent biases, and explores solutions for gender and racial biases in AI on an ongoing basis. Through the proposed collection and analysis of secondary and primary data as well as discussion of implementing proposed solutions, this paper will suggest methods to reduce bias in emotion AI.

Keywords: emotion AI, ethics of emotion AI, privacy and security, gender and racial bias in AI, business

INTRODUCTION

Advancing technology is producing smarter and smarter devices and helping businesses run smoother and faster in a more meaningful way. The next generation of computers may be able to effectively determine how the user is feeling and adapt accordingly. The emerging technology of emotion artificial intelligence and emotional machines is what will make this possible to a large extent. This new possibility brings with it a whole new set of potential (Buolamwini, 2019; Feast, 2019; Tinsley and Ely, 2018; Wellner and Rothman, 2019). The capabilities of devices with emotion artificial intelligence (AI) affords significant advantages for businesses to understand their customers and manage their employees. These same advantages afforded to managers can translate to disadvantages for the customers and employees. The ethics of emotion AI is yet to be explored. Critical questions such as 'what ethical issues does emotion AI present?' and 'how will these issues be addressed?' are yet to be answered. Before such questions can be answered, the capabilities and implications of emotion AI need to be understood more fully. This paper aims to facilitate better understanding of emotion AI and its limitations by gathering and analyzing information from past studies, then examining research gaps, and developing proposed methodology and processes for further research.

Emotion AI may seem like a relatively new development. However, in 1955 MIT media lab professor Rosalind Picard published a study on affective computing, or the ability of computers to recognize human emotion (Somers, 2019). The term 'artificial intelligence' was created that same year although the concept of an intelligent machine had been present in scholarly publications and literature as early as 1308 (Press, 2016). With the ability to identify human emotions, computers could better apply machine learning to interact smoothly with individuals (Banafa, 2016). Emotion AI operates by using emotional indicators, facialexpressions being a main indicator, to determine and react to the emotional state of the user (Kleber, 2018).

Bernard Marr in his Forbes article ‘What is Affective Computing And How Could Emotional Machines Change Our Lives?’ raises the question "Why do we want a computer to empathize with us?" (2016). Marr's answer is simple; emotion is huge part of life for humans therefore computers, which we have come to rely more and more on, are missing a large part of interactions when they cannot identify these emotions (2016). Devices capable of identifying and reacting to human emotions will enhance human and artificial intelligence interactions to become more seamless. The typical AI experience will not feel as one sided as they currently can be. If a computer can recognize and react accordingly to emotions, it is no longer missing as much of the verbal and nonverbals cues in human communication (What is Emotion AI?, 2019). An AI that is aware of emotions can respond to them and potentially imitate them as well (What is Emotion AI?, 2019). The possibilities extend through many fields from healthcare and educational systems, to business, specifically management and marketing (Somers, 2019). With the insight of emotion AI, managers can single out better suited candidates to hire as well as monitor their job satisfaction, reducing costly employee turnover (Moore, 2018). Happy employees translate to happy customers, who can be further served by emotion AI in customer service (Moore, 2018).

The combination of emotion artificial intelligence and marketing creates obvious potential for companies. Because of this, research into emotion AI began with market research and marketing applications (Moore, 2018). Marketing can utilize emotion AI to determine how a viewer feels about a particular advertisement or product (Truthify brings out emotion AI app, 2018). Predicting how a market will react to a new item would be more accurate than ever, and companies can even better understand how to market their products to various audiences. With the internet today, advertising can be personalized for the user, but emotion AI takes this a step further to become even more tailored to the individual and therefore more effective.

Small businesses in particular stand to benefit from the rise of emotion AI. Artificial intelligences already bring a whole host of skills and information to the fingertips of small business owners who often struggle from the lack of specialized knowledge available to sole proprietors (Mills, 2019). With emotion AI, small business owners can have a personal assistant with the ability to vet new applicants and aid in customer interactions, in addition to keeping track of finances (Mills, 2019; MacFarland, 2019). Large companies have already begun implementing emotion AI and those same benefits can be applied to small businesses. Emotion AI can make the hiring process faster, easier, and more reliable (Dattner et al., 2019), an assistance local companies with high employee turnover will certainly appreciate. Call centers have already applied emotion AI to both interact with customers and help agents interact with customers and small businesses can too (Moore, 2018; McFarland, 2019). Emotion AI can automatically send thank you and follow-up messages to customers after a purchase, leaving one less routine task that owners, managers, and employees have to remember (Small Business in the Age of AI, 2018). While 66 percent of small businesses have implemented automation technology, only 11 percent are using artificial intelligence (Small Business in the Age of AI, 2018). Emotion AI has much to offer small businesses but has yet to be applied to its full potential.

Increased effectiveness brings with it increasing concerns over privacy. The very same advantage that marketing gains from machines that can read emotions can be a disadvantage to the customers. Emotion AI provides companies with more information than ever on its target markets while the target market gains another layer of information collection to worry about. It has yet to be determined how and where, if at all, users give permission for their devices to track their emotions (Kleber, 2018). Users might also be concerned about to what degree they give permissions for their devices to monitor their emotional state. Effective emotion AI would rely on visual and audio for greater accuracy, to which users might give permission separately according to personal preference, or have to approve both or neither. Once permissions have been granted, the privacy of the users comes into consideration. In the past, users' identifications have been kept private in data collection through the processes of data anonymization and de- identification (De Montjoye et al., 2017). Data anonymization uses a series of processes to separate the identity of the user from the data (De Montjoye et al., 2017). De-identification is designed to prevent the data from being re-associated with the user (De Montjoye et al., 2017). However, modern devices collect and attach more data to users than ever before, making these two processes more difficult and less effective than before (De Montjoye et al., 2017). The idea behind data anonymization and de-identification is "If the data cannot be associated with the individual to whom is relates, it cannot harm that person," (De Montjoye et al., 2017). Ideally, the complete separation between users and data could be achieved. However, this idea is becoming more difficult to achieve and may even prove impossible. In the more likely case that users cannot be disassociated from the data, and an AI interaction results in harm, the line of accountability is hazy (Keng and Wang, 2018). It is unclear who bears the responsibility for the actions of an artificial intelligence and the repercussions that follow. With the growth of emotion AI, a whole new level of data collection is added, increasing these privacy concerns and the potential for misuse.

A portion of the negative potential of misuse for emotion AI actually arises from the artificial intelligence itself. Emotion AI has developed gender and racial biases. Facial recognition software used in identifying emotional indicators has shown higher error rates for women and individuals of color compared to light-skinned young men (Wellner and Rothman, 2019). AI hiring programs have demonstrated preference for resumes from men over women when selecting candidates for interviews (Lewis, 2018). In recognition of this problem, various studies have identified solutions, such as expanding training datasets, removing gender from data, and implementing anti-bias algorithms (Buolamwini, 2019; Feast, 2019; Tinsley and Ely, 2018; Wellner and Rothman, 2019). As of yet, these solutions are mainly proposed solutions that are not fully tested. When considering possible solutions to problem of bias in emotion AI, the following potential research questions emerge:

Research Question 1: Can expanding training datasets reduce gender and racial bias in emotion AI to a reasonable extent?

Research Question 2: Can doing away with the gender variable singularly and unequivocally from real-time data eliminate gender bias in emotion AI to a reasonable extent?

Research Question 3: Can implementing anti-bias algorithms reduce gender and racial bias in emotion AI to a reasonable extent?

Each of the proposed solutions are promising. However, they may produce their own complications and issues. The remainder of this paper seeks to understand how bias arises and how effective the above proposed solutions may be.

LITERATURE REVIEW

Background Work

The technology of emotion AI is not an entirely new field. The concept has been explored for over fifty years and companies are already applying related programs in many different areas, from retailing to banking to health sciences to teaching (Moore, 2018; Somers, 2019). While emotion AI's capabilities are still rather limited, some researchers have already begun to examine the ethical implications of emotion AI. There is disagreement over how pressing the matter is presently. Some claim there is no hurry, as current AI needs considerable improvements before it will be reach its promised potential, while other claim that ethical problems should be considered now because when emotion AI reaches that point, it will spread fast, leaving researchers playing catch-up (Keng and Wang, 2018). As it is, managers have already begun to implement AI in the decision-making processes. Researchers question what impact this has on management ethically and legally in deciding how accountability comes into practice (Khalil, 1993). Even if the capabilities exist for emotion AI to take a significant role in management, it must be considered to what extent emotion AI should actually be implemented in the decision-making process (Khalil, 1993).

Defining Emotion AI and Ethics

Emotion AI allows the use of artificial intelligence to recognize emotions by verbal and nonverbal communications such as facial expressions (Telford, 2019). AI can compare reactions to situations and the context that created them, then use that data to recognize the same emotions in other people (What is Emotion AI?, 2019). By compiling the reactions of different people to the same image or video, artificial intelligence (AI) can identify emotions and how they are conveyed in facial expressions (What is Emotion AI?, 2019). A machine that can read emotions can then react to them, opening up possibilities for many different fields (Kleber, 2018). Since computers can identify miniscule changes in expression that humans would miss, theoretically emotion AI would be far more effective at reading emotions than humans (Telford, 2019). The potential of emotion AI raises a number of ethical questions. Can an artificial intelligence express bias and, if so, how will that be corrected? Is the use of emotion AI, particularly in certain fields like marketing and advertising, manipulation? Emotion AI can help managers and employees alike but there are ethical issues that need to be addressed first (Whelan et al., 2018).

Current Capabilities of Emotion AI

Before researchers can begin to address these and similar ethical issues, they will need to know what emotion AI is currently capable of and what its limitations are. Devices that can read emotions promise potential in many different fields, including but not limited to management, hiring, retail, and advertising (Moore, 2018; Somers, 2019). With the insight of emotion AI, managers can single out better suited candidates to hire as well as monitor their job satisfaction, reducing costly employee turnover (Moore, 2018). Happy employees translate to happy customers, who can be further served by emotion AI in customer service (Moore, 2018). By monitoring employees' emotional states, AI could prevent costly mistakes by distracted or stressed individuals (Whelan et al., 2018). Emotion AI promises to be even more effective than conventional methods of monitoring, like surveys, because often people do not even realize how they are feeling while a computer can pick up on subtle and overlooked indicators. This combination of abilities implies that computers could become better and faster at reading human emotions than human beings themselves (Whelan et al., 2018; Somers, 2019).

It is important to note that these advantages are not limited to large firms with sizable funds but extend to small businesses as well. In fact, small businesses stand to benefit substantially from emotion AI. Small businesses can struggle with attracting and retaining quality employees. Computers that can read emotions would help with high employee turnover, as well as aid in engaging customers, improving sales, and streamlining the decision-making process (McFarland, 2019). As small businesses can suffer from a lack of access to specialized knowledge due to their size, researchers already foresee emotion AI fulfilling significant roles in companies with few employees (Mill, 2019).

However, these promising capabilities come with limitations. Individuals do not always make the expected facial expressions in response to a situation. The disconnect causes problems for the development of emotion AI (Telford, 2019). For instance, people smile when they are happy but also when embarrassed, pained or uncomfortable. A study has shown that there are nineteen different types of smiles, only six of which are connect to positive feelings (Purdy et al., 2019). Nor do people always vocalize what they are actually thinking, (Purdy et al., 2019). Even if everyone made the expected facial expressions, those expressions associated with a particular emotion vary culturally and geographically, which emotion AI currently lacks the ability to differentiate (Purdy et al., 2019; Telford, 2019). There are other indicators of emotion, such as body language, tone, and word choice, but present emotion AI devices only look at facial expressions (Telford, 2019). Indicators in a person’s voice and word choice express emotion beyond what facial expressions might show, meaning emotion AI does not always have to rely on cameras to determine the emotional state of the user (What is Emotion AI?, 2019). Artificial intelligence that can recognize voice inflections and associate them with the appropriate emotions are more effective so attention to those factors increases accuracy (Somers, 2019).

However, even a combination of all of these indicators does not necessarily guarantee 100% accuracy. The context of a situation also influences emotions and must be taken into consideration when determining how a person feels (Telford, 2019). How one individual reacts in a situation may not be exactly how another individual will react in the same situation.

Additionally, in order to accurately determine emotions, computers will need a baseline with which to compare new data. Attempts at using emotion AI by TSA to identify potential terrorists is not sophisticated enough to lead to active prevention of terrorism through arrests. Even though the emotion AI used by TSA analyzed behavior in addition to facial expressions, establishing a baseline for comparison proved difficult (Telford, 2019). Finding this baseline for any application of emotion AI is proving problematic. 50% of the time, people do not make the expected facial expressions in a situation and the facial expressions associated with a specific emotion differ between cultures and parts of the world (Telford, 2019; Somers, 2019). How each individual expresses emotion facially differs based on factors that include culture. Emotion AI's will need to the abilities to identify a significantly wider range of indicators. Adapting emotion AI's to recognize all the same indicators of emotion that humans see will not solve this problem as effectively as might be hoped. A machine that sees the same indicators as a human will be as good and just as bad at recognizing emotions as human beings are (Telford, 2019). There are many other factors that have yet to be considered. Varying levels of engagement from the user alters results (Purdy et al., 2019). Facial recognition has struggled with varying skin tones and ages such that cameras find it more difficult to recognize emotions and determine facial.

Even with these limitations, emotion AI is already being put into use today. Marketers utilize emotion AI to determine how a customer feels about a product or advertisement, and react accordingly (Somers, 2019). Customer service centers are using the technology to identify angry customers and direct them to human agents instead of automated agents (Kleber, 2018). In 2009, Philips watch company partnered with a Dutch bank to develop a watch that would alert traders when they are stressed or caught up in the 'trading frenzy' so that traders can confirm their rationale before making a trade on impulse (Kleber, 2018). The company Brain Power has produced glasses that to help autistic individuals to recognize emotional tells in other people.

China is already implementing AI in classrooms to measure student learning, although with varying degrees of success due to differing learning styles and levels of engagement (Purdy et al., 2019).

Emotion AI's capabilities and potential, even with its current limitations, has reached a height that affords it separate conferences like Intelligent Virtual Agents and journals like IEEE Transactions (Perepelkina and Vinciarelli, 2019). Studies and discussions in both areas have collectively discovered that people react generally the same to machines as they react to other people (Perepelkina and Vinciarelli, 2019). In fact, a study from the International Journal of Management shows a 75% approval rating for AI devices in customer service and 85% of marketing programs that plan on implementing emotion AI devices as the technology develops (Gursoy et al, 2019). A combination of willingness and intention determines a customer's acceptance of an AI device in the service industry (Gursoy et al., 2019). Disapproval results when actual evaluation differs from expectations and cognitive dissonance causes discomfort for the customer (Gursoy et al., 2019). Another source of discomfort as emerged as researcher explore the ethics of a machine that can read and react to emotions.

Ethical Implications of Emotion AI

Emotion AI's potential in the hiring process provides a good example of the ethical questions that arise. The use of computers capable of reading interviewee's emotional state promises to make hiring faster, easier, and more reliable (Dattner et al., 2019). However, it has yet to be considered if the use of emotion AI in hiring is accurate, legal, and ethical. There are personal questions that employers cannot legally ask but an AI's algorithms could determine the answers indirectly (Dattner et al., 2019). Similarly, marketers can use emotion AI to measure a viewer's response to an advertisement and adapt their marketing strategy accordingly (Truthify brings out emotion AI app, 2018). The data collected and used by artificial intelligences can potentially be misused, raising privacy concerns (De Montjoye et al., 2017). This leads to the question of whether the use of emotion AI will be an opt-in or opt-out type of program and if permission can be revoked at any time without consequences (Kleber, 2018). In the past, methods could be applied to protect the identities of the individuals from whom the data was collected. Data anonymization removes identifying factors from data under the idea that if the individual from whom the data originated could no longer be identified, the data cannot be used to harm them (De Montjoye et al., 2017). De-identification generalizes data to the extent that too many individuals could have created the data such that pinpointing the originator is difficult if not impossible (De Montjoye et al., 2017). These two processes are no longer more effective today than they used to be before, such that technology nowadays is collecting more data and more context is attached to that data than ever before (De Montjoye et al., 2017). Researchers have begun to reasonably hypothesize that it would be next to impossible in the near future (at least from the perspectives of data points accessed) to completely anonymize data (De Montjoye et al., 2017). In the case that users cannot be disassociated from the data and an AI interaction results in harm, the line of accountability is hazy (Keng and Wang, 2018). With the growth of emotion AI, a whole new level of data collection is added, increasing these privacy concerns and the potential for misuse (De Montjoye et al., 2017). Researchers will also have to consider what might happen should an emotion AI fail. Failure in certain situations, or simply flat out, could have long term consequences (Kleber, 2018).

Emotion AI is predicted to have positive impacts on marketing but could lead to dependencies with potential repercussions. With an AI to monitor emotional well-being and react accordingly, companies will be able to better predict and meet the demands of their customers.

This begs the question of what will happen when an emotion AI is unexpectedly inaccurate. A company might invest millions on a product based on information provided by skewed data results. Computers are generally thought of to be impartial, but creators might unwittingly pass along their own personal preferences or biases into the programming of an artificial intelligence (Keng & Weng, 2018). For example, an AI might have in its programming a code that tells it that little girls like pink and consequently fails to play advertisements for a pink sweatshirt to adults, cutting out a portion of the potential market for that product.

Biases in Emotion AI

Examples of bias can be much more severe than the pink sweatshirt example. Gender and racial biases already existed in software and is only aggravated by artificial intelligences (Wellner and Rothman, 2019). Bias can originate with linguistics (Feast, 2019). Certain words are considered masculine or refer to typically male roles, others female and refer to female roles, for example the words doctor and nurse respectively. The AI takes those words and the programmed associations at face value and expands from there, creating a gender bias that surprised researchers and programmers alike (Wellner and Rothman, 2019). Alternately, bias can arise from training techniques (Feast, 2019). AI facial recognition is less accurate with women, older individuals and those of dark skin tones than with light skin-toned, young men (Wellner and Rothman, 2019).

Global companies like Amazon have already tried and abandoned AI hiring programs because of this bias (Lewis, 2018). When diverse workforces have been shown to perform better, hiring processes that automatically dismiss qualified candidates based on gender or race are alarming, both for large companies that focus on sustained growth and for small businesses that already struggle to hire and retain quality employees (Lewis, 2018; MacFarland, 2019).

On the other side of the spectrum, artificial intelligences have helped researchers recognize biases when it comes to emotion but may prove to extend these same biases. A study published by the National Academy of Science has used AI to examine emotion stereotypes related to gender and race over the past century (Garg et al., 2018). From analyzing texts from the 1910s to 1990s, certain words have lost their gender and racial associations, while others have not changed. 'Resourceful' and 'clever' are no longer as masculine as they once were, but words like 'alluring' and 'homely' have become no less feminine. Some of these changes are related to race, such as terms for outsider are not as strongly associated with Asian names (Garg et al., 2018). These reductions in gender associations with words are promising, especially since linguistics has been shown to influence biases in artificial intelligences. However, this may not be enough as emotion AI grows more common. Emotional stereotypes based on gender and race may end up aggravated by the technology. The most common stereotypes in this regard are for women to be over-emotional and men to be under-emotional (Durik et al., 2006). No scientific evidence has been found to support this, however the word 'emotional' was more commonly found referring to women than men (Heesacker et al., 1999; Hutson, 2018). When these stereotypes are applied in the workplace, women are often considered less capable of leadership positions, causing AI hiring systems to discount resumes submitted by women for management roles (Lee et al., 2019). Similar stereotypes expect African American women to display more interest, Hispanic American men to express more pride and shame, men in general to show more anger, and women in general to show more embarrassment, fear and sadness (Durik et al., 2006). Artificial intelligences have already been shown to pick up on human biases and apply them widely in expected scenarios (Lee et al., 2019; Wellner and Rothman, 2019). With the ability to identify emotions and therefore emotion stereotypes, emotion AI could further this predicament (Purdy et al., 2019).

Because of the problems related to bias and the effectiveness of emotion AI in general, managers and programmers need to be aware of potential mistakes and biases of computer systems, perhaps especially emotion AI. They need to continually monitor and correct problems (Khalil, 1993). Awareness of the potential for issues would lead them to seek solutions.

Solutions that might include expanding training datasets, entirely removing gender from datasets, transparent algorithms and anti-bias algorithms (Lee, 2018; Wellner and Rothman, 2019).

DISCUSSION AND SYNTHESIS OF INFORMATION

Research Gaps

As an emerging technology, there is still have much about emotion AI that is yet to be fully understood. The previously discussed potentials are only the start of the technology's capabilities. These promising capabilities in turn raise concerns about the uses and impact of emotion AI. What effects emotion AI might have on the individual users has yet to be fully explored. Privacy, security, and ethics are questions to be asked particularly about the combination of emotion AI and marketing. Marketers already seek to evoke specific emotions from viewers with advertisements. They use what they know about their target market's feelings towards certain products to craft an effective strategy. The effectiveness increases if marketers knew what emotions each individual viewer is experiencing. Emotion AI could glean personal details from a viewer's facial expressions as they react to a product. The advertisement goes from tailored to a target market to tailored to a specific individual. The question has already been raised if the use of emotion AI in fields like marketing and advertising is manipulative and unethical (Whelan et al., 2018). The act of collecting data on how a person feels is collecting a more intimate level of information about that individual. Researchers will have to consider under what circumstances, if any, the use of emotion AI to influence a viewer to buy a product or support an organization is acceptable and ethical.

Another topic of concern, bias, similarly raises crucial questions about the ethics of emotion AI, particularly the combination of emotion AI and business. Researchers have already found that facial recognition and emotion AI struggle to accurately identify faces and expressions of darker skin tones. Programmers will need to adapt artificial intelligences to recognize far more diverse individuals in order to be at all effective on the broad scale it will face in the real world outside of the lab.

Theoretical Perspectives

Various potential solutions exist for these issues listed previously. Privacy concerns revolve around laying down clear rules and permissions for users and businesses. To accommodate this, laws will likely have to be extended or adapted to include privacy with the use of emotion AI devices, as has been done for over time for each new wave of technology. Other issues require more work on the part of the programmers to resolve. Emotion AI has already shown inherent biases also revealed in pre-existing software (Wellner and Rothman, 2018). Removing gender completely from datasets has been proposed to effectively eliminate gender bias in emotion AI (Wellner and Rothman, 2019). AI hiring software like Pymetrics and GapJumpers use skills-based tests, rather than resumes containing information on gender and race, to select candidates for interviews and the resulting interviewees represent a far more diverse background (USA staffing services: Eliminate unconscious bias in recruiting: Choosing the right AI solution, 2020). While this seems like the ideal solution, removing gender completely may cause trouble in issues where gender does need to be considered. For example, women's maternity leave ended up counting against them in AI performance review programs that ranked employees for promotions (Tinsley and Ely, 2018). AI's have also been found to use other identifying information, such as zip codes, to discriminate based on gender and race even when specific data is excluded. Since emotion AI exaggerated gender and racial biases already present in computer programs (Wellner and Rothman, 2019), the same could occur for emotion AI with similar purposes like the performance reviews.

As an alternative or addition to removing gender from training datasets, researchers recommend expanding datasets. Programmers could include diverse pictures when developing facial recognition so that the AI can accurately identify expressions on darker skin tones, which artificial intelligences have been noted to struggle with currently because training datasets tend to be of primarily light skin tones (Wellner and Rothman, 2019). Training datasets currently contain mainly light-skinned, young men, causing emotion AIs to struggle identifying emotions in women, individuals of color, and the elderly (Feast, 2019). The error rates remain low for young, white men, at about 1% error, while women of color experienced 35% error (Buolamwini, 2019). Expanding the diversity of training datasets will teach AI's to accurately identify a diverse range of individuals. However, this only addresses AI bias in facial recognition.

A common theme in these solutions to bias in emotion AI is to focus on the problem at the source and be aware of the potential for bias from the beginning (Wellner and Rothman, 2019). For this reason, Wellner and Rothman propose transparent algorithms (2019). If programmers can identify how the AI reached the decision, it can identify how the bias arose and correct it. However, not all researchers are confident this will fully reveal how the biases occur (Wellner and Rothman, 2019). Instead, anti-bias algorithms have been suggested. Rather than attempt to create algorithms that avoid bias from the beginning, thus trying to predict outcomes that have already proven surprising to programmers and researchers alike, create algorithms that specifically address the issues as they arise (Wellner and Rothman, 2019). This solution holds potential when considering the many different ways bias develops in AI.

Recent studies into emotion AI have not only identified such areas for concern but also have begun proposing solutions to these issues (Buolamwini, 2019; Feast, 2019; Khalil, 1993; Tinsley and Ely, 2018; Wellner and Rothman, 2019). However, the full potential and possible drawbacks of these proposed solutions have yet to be fully explored.

CURRENT RESEARCH INTO EMOTION AI

Emotion AI promises potential for many different fields, some more than others. Fields that stand to benefit especially from machines that can read emotions include education, medicine, automotive, marketing and management. As such, often studies focus on these fields (Khalil, 1993; Kleber, 2018; Lewis, 2019). Because of the obvious potential the combination of emotion AI and advertising affords, companies developing that technology generally begin with the marketing department (Moore, 2018). Other studies seek to improve emotion AI's capabilities in general (Barrett et al., 2019; Bartneck, Lyons, and Saerbeck, 2017; Kaiser, 1994; Pierre-Yves, 2003).

Many of these studies seeking to generally improve emotion AI have agreed that there is more to human communication than simply words and speech. Nonverbals play a crucial part of communication (Kaiser and Wehrle, 1994). Even while accepting that nonverbals are crucial to communication, initial emotion AI prototypes have relied mainly on only one form of nonverbal, facial expressions. Identifying emotions from facial expressions alone has proven difficult enough. Variances in cultural and geographical indicators (Telford, 2019) and variances in meanings for a single expression (Purdy et al., 2019) make isolating the intended emotion problematic. To make the task harder still, recent emotion AI studies have overturned the assumption that emotions can easily be determined from facial expressions (Barrett et al., 2019). A study at the University of Geneva in Switzerland struggled to determine how to measure faces so that facial expressions in turn could be measured but found variety in faces and individual expressions make establishing a standard for comparison difficult (Kaiser and Wehrle, 1994).

Researchers in another study built their research around the idea that a facial expression is actually a combination of facial movements and the differing combinations of facial movements create expressions with different meanings (Barrett et al., 2019). The University of Geneva study proposed a solution of digitalizing faces for easier computer analysis and computer simulation (Kaiser and Wehrle, 1994).

While these studies have examined the standards and methods for universally identifying human emotions, other studies focus on how to adapt artificial intelligences to mimic emotion. An article in the International Journal of Human-Computer Studies discusses the development of algorithms that will allow robots to imitate emotions through tone of voice (Pierre-Yves, 2003). Current AI voices with intonation are adult and speak mostly English (Pierre-Yves, 2003).

Theoretically, computers that can both recognize and imitate human emotions would further streamline human and artificial intelligence interactions. However, an experiment at the University of Duisberg-Essen in Germany discovered potentially unsettling results when artificial intelligences displayed emotion (Parry, 2018). Study participants struggled to follow through switching off a cute, little robot named Nao when Nao begged not to be turned off at the end of the conversation (Parry, 2018). The difficulty occurred both when the robot exhibited more human-like speech features, such as intonation and friendly comments about pizza, and when the robot spoke in a monotone associated with computers and appliances (Parry, 2018).

When Nao exhibited human-like behaviors, participants experienced more anxiety over the decision when it came time to turn the robot off. When Nao spoke monotone, the sudden display of emotion, asking not to be turned off, caused participants to deliberate longer before making a decision (Parry, 2018). Some participants refused flat out to turn off Nao because the little robot begged them not to (Parry, 2018). The results raise concern over the potential for manipulation and misuse (Parry, 2018). It has already been questioned if the use of emotion AI in advertising and marketing is manipulative and unethical (Whelan et al., 2018). Adding the ability to mimic emotions to artificial intelligences would only increase the potential for manipulation.

With this potential in mind, Andrew McStay, a professor of digital media at Prifygol Bangor University, spoke to 2,000 United Kingdom citizens about the use of AI in advertising (Lewis, 2019). McStay informed participants during November of 2015 that cameras in Piccadilly Circus scanned the crowd looking for indicators, such as facial expressions, to determine which advertisements to show on the screens (Lewis, 2019). Over half the citizens he spoke to expressed discomfort over the idea that cameras were watching them and judging their reactions for advertising purposes. About one third said they did not mind, as long as no identifying information was recorded. McStay found younger individuals were more likely to not mind, while those over the age of 65 expressed more suspicious of being unknowingly monitored (Lewis, 2019).

MIT graduate Joy Buolamwini discovered another reason to be suspicious of artificial intelligence in 2015 (Buolamwini, 2019). Because training datasets consisted mainly of young, white males, facial recognition software failed to identify her darker skin tones. She dug deeper into the problem to find that while light skinned men experienced about 1% error, women of color experienced 35% error (Buolamwini, 2019). AI facial recognition error increases with women and individuals with dark skin tones (Wellner and Rothman, 2019). The discovery prompted her to create the Algorithmic Justice League to raise awareness of this issue. Previous studies discovered that gender and racial biases already existed in software and are only aggravated by artificial intelligences (Wellner and Rothman, 2019).

The aggravation of biases by artificial intelligences is explored in a research article by Galit Wellner and Tiran Rothman (2019). In the article, Wellner and Rothman (2019) discuss a couple of contributing factors leading to gender and racial biases in AI. The first key factor is linguistics. Language contains ambiguities and connotations that computers took at face value, as they had no code in their programming that would tell them to adjust for ambiguity the way a human being automatically would. Words like doctor and nurse, which have masculine and female associations respectively, implied gender roles that AI's carried over into other areas, such as hiring processes (Wellner and Rothman, 2019). Wellner and Rothman (2019), just like Buolamwini (2019) discovered, also discussed the disproportionate rates of error for men, women, and individuals of color in facial recognition software.

Over the course of the studies introduced above, a general consensus emerges. Emotion AI currently lacks the ability to accurately and reliably identify emotions due to facial variances. Combining artificial intelligence and the capabilities to recognize and imitate emotions increased the potential for manipulation and misuse. Emotion AI currently exaggerates gender and racial biases already inherent in artificial intelligences. Finally, individuals are generally uncomfortable with idea of artificial intelligences watching them, and understandably so considering the previously mentioned potentials and biases. To eliminate, or at least greatly reduce bias and the potential for misuse, programmers and managers will have to develop a standard for facial analysis that recognizes diversity as well as establish privacy and security policies that will protect individuals.

PROPOSED FUTURE RESEARCH METHODOLOGY AND ANALYSIS

Research Design, Methods, and Model Diagram

Past studies on emotion AI has revealed areas of concern requiring more research, although emotion AI yet lacks the ability to fulfill these potential issues. Innovators are currently working on creating artificial intelligences that are increasingly better at reading and imitating emotions. Machines with the capabilities to understand and possibly mimic them demonstrate new opportunities for manipulation and misuse, which in the case of inherent AI biases can occur without the creators' intent or knowledge. As shown in the above studies, artificial intelligences can be put to broad uses, of individuals are already uncomfortable. This feeling of mistrust is justified when AI's inability to accurately identify emotions and inherent biases are considered.

Combining emotion AI's broad potential uses with its failure to remain neutral, as many would expect of computers and machines, could result in disconcerting and potentially disastrous effects. Bias takes shape in different ways involving emotion AI, starting with its inability to achieve high accuracy in identify emotions of even a typical dataset. Artificial intelligences have displayed gender and racial biases even before emotional indicators and imitation were added.

These biases develop in different ways and heighten each other's effect. For example, connotations and ambiguities of language are taken as givens by computers and are then reapplied in other scenarios (Feast, 2019). The results startle programmers and researchers (Wellner and Rothman, 2019).

To begin correcting such behavior, researchers will need to better understand how biases form in artificial intelligences, how artificial intelligences use the information they collect, and what corrective methods are available. Previous studies generally agree that training datasets are main contributors to the issue. Wellner and Rothman (2019) have suggested removing gender entirely from datasets. While this seems like the ideal solution, removing gender completely may causes problems in situations where gender does need to be considered. For example, women's maternity leave ended up counting against them in AI performance review programs that ranked employees for promotions (Tinsley and Ely, 2018). Research will need to be conducted that explores what happens when gender is removed from datasets, if it can reduce or eliminate gender bias problems in emotion AI and under what circumstances this does or does not occur.

This can be achieved by a few different recommended methods (see Table 1).

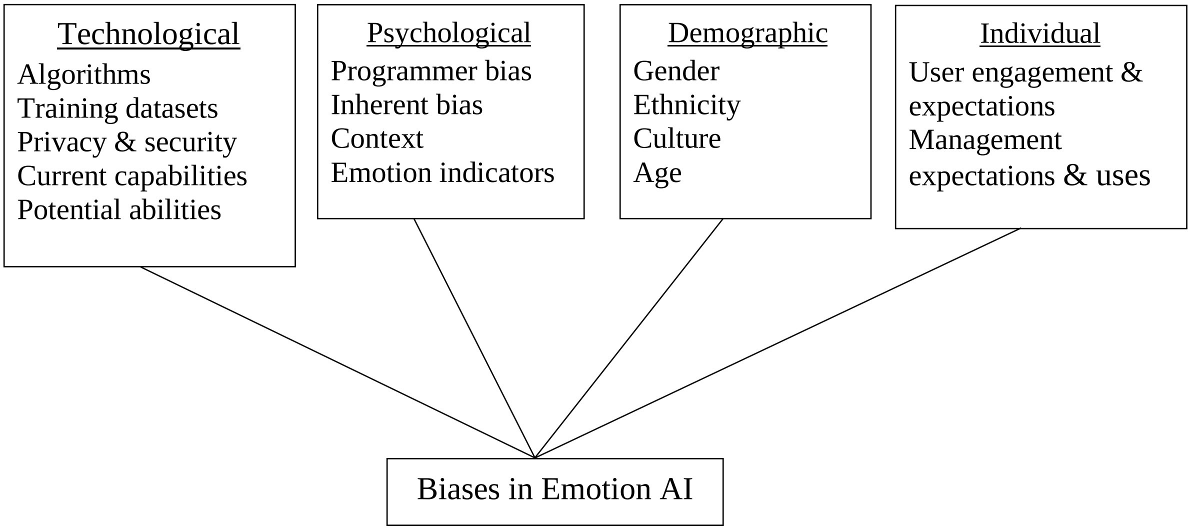
TABLE 1
PROPOSED RESEARCH METHODS

Research Method	Description	Reasoning
Secondary data	Collecting data from past studies on emotion AI in business and ethics	Researchers can gather data from previous studies to guide new research

Observation	Watching and recording how the AI acts in various scenarios	Allows researchers to observe potential problems and trace how they might have occurred
Experiment	Altering training datasets and algorithms	Programmers can alter variables of emotion AI and view the results
Focus Groups	Allowing participants to interact with the emotion AI and interviewing them before and after	Researchers can benefit from outside input on alterations to emotion AI

Variables, such as training datasets, related to emotion AI will have to be considered over the course of research to effectively eliminate gender and racial biases (see Figure 1). Past studies on gender and racial bias in emotion AI have already agreed that training datasets will need to be diversified in order for artificial intelligences to be capable of more accurately recognize emotions in individuals (Buolamwini, 2019; Feast, 2019; Tinsley and Ely, 2018; Wellner and Rothman, 2019). Given that AI's will be analyzing a wider range of individuals in a diversified training dataset, these datasets may need to be expanded in number so that the computers can more easily discern and trace trends.

FIGURE 1
RESEARCH MODEL DIAGRAM



CONCLUDING REMARKS

Proposed solutions to reducing racial bias include expanding training datasets to increase diversity and adding anti-bias algorithms. The effectiveness of these has yet to be tested.

Solutions to gender bias from previous studies suggest removing gender entirely from training datasets (Wellner and Rothman, 2019). This is predicted to eliminate gender bias, however, it may cause problems in cases where gender ought to be considered. Even still, removing gender completely from training datasets should reduce gender bias. Other solutions include diversifying training datasets and programming anti-bias algorithms but have yet to be fully explored.

As emotion AI becomes more effective, growing concerns will have to be addressed and hopefully resolved without lessening the advantages the technology could afford. Each successive generation of computers and devices brings with it increased capabilities, penetrating further into everyday life. Machines that can read emotions promise potential for many fields, business in particular. With the capabilities to recognize and react to human emotions, possibly even imitate them, emotion AI could revolutionize many aspects of business and daily activities. The rise of emotion AI raises its own set of ethical complications, including limited capabilities, gender and racial bias. Bias in AI is not an entirely new phenomenon yet solutions to these issues have not been fully tested. Reconciling possibilities and concerns of emotion AI requires further extensive research on an ongoing basis.

Declaration of Conflicting Interests

The authors declare that there has been no conflict of interest whatsoever thereon and thereof.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

Banafa, A. (2016). What is Affective Computing? OpenMindBBVA. Retrieved from https://www.bbva openmind.com/en/technology/digital-world/what-is-affective-computing/

Barrett, L.F., Adolphs, R., Marsell, S., Martinez, A.M., & Pollak, S.D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Association for Psychological Science, 20*(1), 1–68. Retrieved from https://journals.sagepub.com/stoken/default+domain/10.1177%2F15291 00619832930-FREE/pdf

Bartneck, C., Lyons, M., & Saerbeck, M. (2017). *The Relationship Between Emotion Models and Artificial Intelligence*. Eindhoven University of Technology. Retrieved from https://arxiv.org/ pdf/1706.09554.pdf

Buolamwini, J. (2019). Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It. *Time Magazine Online*. Retrieved from https://time.com/5520558/artificial- intelligence-racial-gender-bias/

Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The Legal and Ethical Implications of Using AI in Hiring. *The Harvard Business Review*. Retrieved from https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring

De Montjoye, Y., Farzanehfar, A., Hendrickx, J., & Rocher, L. (2017). Solving Artificial Intelligence's Privacy Problem. *Journal of Field Actions*, *10*(17), 80–83. Retrieved from https://journals.openedition.org/factsreports/4494

Durik, A.M., Hyde, J.S., Marks, A.C., Roy, Anaya, D., & Schultz, G. (2006). Ethnicity and Gender Stereotypes of Emotion. *Sex Roles*, *54*, 429–445. https://doiorg.proxy.consortiumlibrary.org/10.1007/s11199-006-9020-4

Feast, J. (2019). 4 Ways to Address Gender Bias in AI. *Harvard Business Review*. Retrieved from https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai

Garg, N., Schienbinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Science*, *115*(16). DOI:10.1073/pnas.1720347115

Heesacker, M., Wester, S.R., Vogel, D.L., Wentzel, J.T., Mejia-Millan, C.M., & Goodholm, C.R. (1999). Gender-Based Emotional Stereotyping. *Journal of Counseling Psychology*, *46*(4), 483–495. Retrieved from http://web.a.ebscohost.com.proxy.consortiumlibrary.org/ehost/detail/detail?vid=0&sid =49b81828-da7b-4307-b47a36a5f617ac3d%40sessionmgr4008& bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=1999-11962-006&db=pdh

Hutson, M. (2018). Artificial intelligence reveals how U.S. stereotypes about women and minorities have changed in the past 100 years. *Science Magazine*. Retrieved from https://www.sciencemag.org/news/2018/04/artificial-intelligence-reveals-how-us-stereotypes-about-women- and-minorities-have

Kaiser, S., & Wehrle, T. (1994). *Emotion research and AI: Some theoretical and technical issues.* Universite de Genve. Retrieved from https://tecfa.unige.ch/perso/wehrle/ OnlineDocs/rai4.pdf

Keng, S., & Wang, W. (2018). *Ethical and Moral Issues of AI: A Case Study on Healthcare Robots*. Americas Conference on Information Systems. New Orleans, LA. ResearchGate. Retrieved from https://www.researchgate.net/publication/325934375 _Ethical_and_Moral_ Issues_with_AI

Khalil, O. (1993). Artificial Decision-Making and Artificial Ethics: A Management Concern. *Journal of Business Ethics*, *12*(4), 313–321. Retrieved from https://www-jstor-org.proxy.consortiumlibrary.org/stable/25072403?pq-origsite=summon&seq=1#metadata_info_tab_contents

Kleber, S. (2018). 3 Ways AI Is Getting More Emotional. *Harvard Business Review*. Retrieved from https://hbr.org/2018/07/3-ways-ai-is-getting-more-emotional

Lee, N. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication & Ethics in Society*, *16*(3), 252–260. doi:http://dx.doi.org.proxy.consortiumlibrary.org/10.1108/JICES-06-2018-0056

Lewis, T. (2019). AI can read your emotions. Should it? The Guardian. Retrieved from https://www.theguardian.com/technology/2019/aug/17/emotion-ai-artificial-intelligence-mood-realeyes- amazon-facebook-emotient

Marr, B. (2016). What is Affective Computing And How Could Emotional Machines Change Our Lives? *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2016/05/13/what-is-affective-computing-and-how-could-emotional-machines-change-our-lives/#72e5b49be580

McFarland, R. (2019). *5 Ways Small Business is Using AI and Machine Learning Right Now*. Cox Blue. Retrieved from https://www.coxblue.com/5-ways-small-business-is-using- ai-and-machine-learning-right-now/

Mills, K. (2019). How AI Could Help Small Businesses. *Harvard Business Review*. Retrieved from https://hbr.org/2019/06/how-ai-could-help-small-businesses

Moore, S. (2018). *Check out these 13 ways emotion artificial intelligence helps companies improve customer experience and unlock cost savings*. Smarter with Gartner. Retrieved from https://www.gartner.com/smarterwithgartner/13-surprising-uses-for-emotion-ai- technology/

Parry, W. (2018). Robot manipulates humans in creepy new experiment. Should we be worried? *NBC News*. Retrieved from https://www.nbcnews.com/mach/science/robot-manipulates-humans-creepy-new-experiment-should-we-be-worried-ncna900361

Pierre-Yves, O. (2003). The production and recognition of emotions in speech: features and algorithms. *Journal of Human-Computer Studies*, *59*(1–2), 157–183. Retrieved from https://www-sciencedirect-com.proxy.consortiumlibrary.org/science/article/pii/ S1071581902001416

Gil Press. (2016). A Very Short History of Artificial Intelligence (AI). *Forbes*. Retrieved from https://www. forbes.com/sites/gilpress/2016/12/30/a-very-short-history-of-artificial-intelligence-ai/#4a25d45f6fba

Purdy, M., Zealley, J., & Maseli, O. (2019). The Risks of Using AI to Interpret Human Emotions. *Harvard Business Review*. Retrieved from https://hbr.org/2019/11/the-risks-of- using-ai-to-interpret-human-emotions

Small Businesses in the Age of AI. (2018). Inuit, SlideShare. Retrieved from https://www.slideshare.net/IntuitInc/small-business-in-the-age-of-ai

Somers, M. (2019). Emotion AI, explained. *MIT Sloan*. Retrieved from https://mitsloan.mit.edu/ideas-made-to- matter/emotion-ai-explained

Telford, T. (2019). ‘Emotion detection’ AI is a $20 billion industry. new research says it can’t do what it claims. *The Washington Post*. Retrieved from https://search-proquestcom.proxy.consortiumlibrary.org/docview/2267296758?pqorigsite=summon&accountid=14473

Tinsley, C., & Ely, R. (2018). What Most People Get Wrong About Men and Women. *Harvard Business Review*. Retrieved from https://hbr.org/2018/05/what-most-people-get-wrong-about-men-and-women

What is Emotion AI? (2019). Behavior Signals Technology. Retrieved from https://behavioralsignals.com/what- is-emotion-ai/

Whelan, E., McDuff, D., Gleasure, R., & Brocke, J.V. (2018). How emotion-sensing technology can reshape the workplace. *MIT Sloan Management Review*, *59*(3), 7–10. http://search.proquest.com.proxy.consortiumlibrary.org/docview/2023991461?accountid=14473

Wellner, G., & Rothman, T. (2019). Feminist AI: Can We Expect Our AI Systems to Become Feminist? *Philosophy & Technology*. Retrieved from http://sz3sa6ce8r.search.serialssolutions.com/?ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=Feminist+AI%3A+Can+We+Expect+Our+AI+Systems+to+Become+Feminist%3F&rft.jtitle=Philosophy+%26+Technology&rft.au=Wellner%2C+Galit&rft.au=Rothman%2C+Tiran&rft.date=2019-05- 18&rft.issn=2210- 5433&rft.eissn=2210-5441&rft_id=info:doi/ 10.1007%2Fs13347-019- 00352-z&rft.externalDBID=n%2Fa&rft.externalDocID =10_1007_s13347_019_00352_z&paramdict=en-US