Request for Information (RFI) on Public and Private Sector Uses of Biometric Technologies: Responses

**DISCLAIMER:** Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government, the Office of Science and Technology Policy (OSTP), or any other Federal agencies or government entities. We bear no responsibility for the accuracy, legality, or content of these responses and the external links included in this document. Additionally, OSTP requested that submissions be limited to 10 pages or less. For submissions that exceeded that length, the posted responses include the components of the response that began before the 10-page limit.
Comments on use of biometric technologies for inference of individual mental and emotional states: Facial Expression Analysis (FEA)

As Artificial Intelligence (AI) increasingly becomes part of everyday life, the opportunities and potential perils posed by using these new technologies are most elevated during the early stages of adoption and implementation. Because the critical decisions concerning how a technology is used affects its development and application, and occur during the early stages of scientific discovery, we welcome the OSTP’s efforts to create a “Bill of Rights for an Automated Society.” More specifically, we embrace the invitation to comment focusing on the opportunities and threats posed by Facial Expression Analysis (FEA).

FEA technology offers to revolutionize the study of human behavior and is thus an exciting tool for social and psychological scientists to add to their toolbox. However, there are multiple evidentiary reasons to proceed with caution, and to consider regulating the use of FEA in hiring, workplace, and education practices; perhaps more importantly, we advise careful consideration of FEA use in the criminal justice system where the powerless have little recourse when technology is misused. The primary point of this comment is that the theory driving how FEA is used and marketed is flawed and threatens to exacerbate current societal inequities. This is especially the case with videoconferencing becoming a ubiquitous part of everyday life and providing digital intimacy via a mediated face-to-face setting (1). In what follows, we provide information requested regarding: RFI-4. Exhibited and potential harms of a particular biometric technology, focusing on 4.a. harms due to questions about the validity of the science used in the system, specifically questions regarding the inference process, and 4.b. harms due to disparities in the effectiveness of the system for different demographic groups. We then discuss policy implications in light of historic management practices and privacy concerns.

Potential harms from Facial Expression Analysis due to validity of inference process

The societal threat posed by the unquestioning acceptance of FEA may be seen as similar to one experienced with the now century-old polygraph. These so called “lie detectors” entered into the popular culture with the expectation that deception could be unerringly identified based upon psycho-physiological indicators being measured and analyzed during questioning. More specifically, respiration, heart rate, blood pressure, and galvanic skin response (GSR - electrodermal response [EDR] in which sweat on the skin’s surface is measured) are used to compare physiological indicators in response to control and the relevant questions (2). Claims regarding the “truth telling abilities” of 90% or better for polygraph lie detection continue to be made regardless of a report by the National Research Council (NRC) in 2003 (2) and an update in 2019 by Iacono and Ben-Shakhar (3) that this assertion greatly overstates its efficacy. To quote the NRC regarding the persistence of the myth of polygraph accuracy “... practitioners have always claimed extremely high levels of accuracy, and these claims have rarely been reflected in empirical
research” (2003, p. 107) with Iacono and Ben-Shakhar concluding: “Fifteen years later, the landscape has not changed: The panel’s conclusions still stand” (2019, p. 96).

With the polygraph still being used in the criminal justice system by state and local law enforcement agencies, in the court system, and by private businesses in the interview process, as well as its continuing cachet as proof of guilt or innocence for political and public figures, the persistent myth of its infallibility is instructive for FEA. While coherence of a person’s narrative with their physiology and resultant nonverbal behavior is instructive in understanding interpersonal trust (4-8), the ability to correlate nonverbal behavior with deceptive statements is weak at best (9). Despite there being a greater connection between truth telling and nonverbal behavior, especially during stressful events (10, 11), many facial behaviors can be controlled (12). This leads to multiple critiques regarding the use of FEA in employment, education, governance, and criminal justice decision making. We organize these critiques based upon three distinct concerns regarding how FEA is used, and subsequently consider how FEA use may negatively affect different populations.

First, the term “Facial Expression Analysis” is a misnomer. The face does not “express” emotions, with internal physiology and possible behavior represented in the face on a one-to-one ratio (13). A meta-analysis by Durán and Fernández-Dols (14) of studies over the past 50 years on coherence of facial behavior with self-reported emotion suggests “that between 6% and 27% of the participants who reported one of the six emotions produced the whole predicted facial expression” (p. 15), with lower estimates for co-occurrence during “an actual emotional situation” (p. 14). In other words, facial behaviors do not occur as predicted; i.e., facial expressions do not consistently express internal physiological states. Likewise, while there is a relationship between facial behavior and emotional experience via facial feedback, a meta-analysis shows this relationship is small and variable (15), potentially due to a range of contextual and individual causes. In sum, a growing corpus of scientific findings have built and elaborated upon the groundbreaking work of emotion science pioneers to provide for a re-examination of “folk” psychological understandings; this has led to a much more nuanced and complex appreciation of facial behavior.

Second, FEA is biased. The major means by which FEA is applied in both academia and the private sector is through the lens of the Basic Emotion Theory which posits six fundamental emotions of happiness, anger, fear, sadness, disgust, and surprise (16). Here, there is a negative bias with four of the six emotions being negative and only one positive (13). Surprise, for its part, is not so much an emotion as a transitory state of orienting oneself to address stimuli having negative or positive implications. Consequently, such diverse and important positive emotions as interest, pride, pleasure, joy (17), and contentment (amongst others) are either bundled into one undifferentiated emotion of happiness or are misidentified. A recent performance comparison of eight commercially available FEA programs by Dupré and colleagues suggests that even with faces posing the six basic emotions, these programs significantly and substantially underperform human observers, with the performance suffering further when considering spontaneous facial behavior (18). The focus on identifying negative emotions in the face encourages regulatory and social strategies involving control and coercion, rather than cooperation and coordination that may result from positive emotions (19).
Third, FEA’s use of Basic Emotion Theory is overly simplistic. Facial behavior itself is highly complex with 44 unique Action Units (AUs) identified by the Facial Action Coding System (FACS) and corresponding with facial musculature, albeit not on a 1:1 ratio. Another 18 Action Descriptors (ADs) describe behaviors that might have several muscles involved with movement; the combination of these AUs and ADs, along with the timing of their appearance, and their comparative intensity influences interpretation (20). For example, the smile serves multiple discrete social purposes, including reflecting amusement, communicating affinity, and asserting dominance (21-24); while all these smiles involve the pulling up of the lip corners using the zygomaticus muscle (AU 12), the involvement of other facial muscles, as well as their movement in terms of intensity and timing, affects how these different smiles are received and responded to. As pointed out by Dupré and colleagues (18) “(B)y extending the number of emotion categories, automated methods might overcome their current limitation of classifying a small set of emotion labels that are insufficient to describe the complexity of human expressive behaviors” (p. 11). Echoing these concerns, modern neurocomputing research remains in the early stages of applying FEA in a natural environment and identifying finer-grained emotional states (25, 26).

In summary, the face may best be seen as a communication tool with which humans influence others both consciously and subconsciously (27-29). It does so through signals that are more-or-less reliable depending on how they reflect internal physiology of an individual (30, 31), with greater reliability through multimodal coherence of verbal utterances and other body signals. For instance, the emotion of amusement may be seen as extremely reliable due to the coherence of multimodal signals beyond the face, with the amusement smile including the zygomaticus muscle being contracted, the jaw dropped, and the eyes closed, to include the vocalics of laughter, and the body movement including shaking in the torso (32). Furthermore, the use of these subtle signals in the face (as well as other nonverbal signals) are affected by contextual elements, including the person with whom the individual interacts. Attempts to simplify and explain behavior, especially for the purpose of gain and/or social control, without accounting for the complexity and variation in human experiences will likely lead to misuse and potentially to exploitation – especially of those social groups with little to no recourse.

Potential harms from Facial Expression Analysis due to demographic group disparities

A major critique beyond the application of simplistic models of emotion, as is currently the case with FEA, is that it simplifies the complexity of human emotional experience both across and within individuals. Such factors as development, gender, and societal context, amongst others, play a role in the display and interpretation of facial behaviors. The simplistic “one size fits all” approach currently in use with FEA technologies for employment, education, service provision, and governance – including criminal justice – may in many cases fail to adequately support local norms and individual freedoms.

Life history influences emotional experiences and the resultant facial behaviors. Research shows that infants develop the ability to process the faces of those around them and appreciate the social context as they grow older, and consequently develop different strategies to influence others (33, 34). This awareness and the deployment of facial behaviors becomes more varied and resonant throughout an individual’s lifespan. This occurs through introspection and interoception, the knowing of one’s own mind and body
states respectively (35) and may even be seen with differences in individual facial behavior repertoires (36).

More broadly, influences at the intersection of biological sex and socially constructed gender affect facial behaviors, with biases being introduced early in life. Gender biases with FEA might occur due to females being more nonverbally communicative than males, and when males are more expressive, having a tendency towards anger and fear (37-39). Tellingly, Woodzicka found that during mock interviews females smiled between half as much to twice as long as males after controlling for time, based upon the type of smile measured (40). Mehu and Dunbar found that gender and power interacted asymmetrically, with younger males deliberately posing smiles to older individuals with (presumed) greater status within a hierarchy. For women the relationship was not as clear. Older women posed more deliberate smiles, while younger women laughed and displayed amusement smiles to a greater extent, suggesting a power dynamic other than hierarchy at work (41). Thus, while men might smile as often, if not more than women do, in more fluid social circumstances (41-43) context and age jointly affect facial behavior in men and women differentially in terms of both types and amount.

FEA might likewise introduce biases against rural citizens due to their being less facially animated and less likely to display positive emotions than those living in more heterogenous urban areas. Cross-cultural research of thirty-two countries shows that greater historical heterogeneity of a culture, i.e., the extent to which numerous source countries contributed that that country's present day population, and residential mobility – the increased likelihood of moving away from one's current home – was positively related to the willingness to openly show emotions (study 1), as well as rationale for smiling (44). This research was followed up with a more extensive world nation-level and state-level polling data (45) finding that "[i]ndividuals from historically heterogenous cultures smile more and display facial expressions that are more accurately recognized across cultures" (p.13).

**Policy Implications**

In the modern era, the application of new technologies within organizations has largely outpaced the ability of the nation’s policymaking system to protect employees – and the general public – from adverse outcomes. Starting in the early 1900s, the “time and motion studies” used by organizations implementing “scientific management” provided managerial tools that relied on the then-new technologies of photography and videography to increase productivity. While such studies allowed for the systematic and scientific discovery of the “one best way” within organizations to accomplish a task, thus maximizing labor efficiency, it also implemented a style of performance management that pressured employees to meet often-unattainable efficiency goals. It also reinforced discriminatory managerial decisions while relying on conceptualizations of organizational behaviors that were often misguided and, at their worst, led to inhumane work conditions and labor strife (46-48). Arguably, negative long-term effects of “scientific management” approaches can be seen in the many organizations that now use monitoring software on a variety of employee behaviors (49).

Given such history and the above risks and limitations of FEA, new concerns will emerge as monitoring software evolves to integrate FEA data and organizations generate inaccurate inferences of individual behaviors and emotions. For instance, organizations
could mis-identify emotional states with FEA and then use that data in subsequent managerial decisions. In such a scenario, FEA software might identify an employee with facial displays commonly associated with anger during a meeting. Retrospectively analyzing FEA data from the meeting, a manager may infer the employee’s emotional state is negative and react with punitive or restrictive measures to control the employee’s behavior in ways that reinforce existing biases and stereotypes, even though FEA software cannot accurately predict an individual’s emotional state. When extended to the criminal justice system, which now depends on evidence-based therapeutic “behavioral change” models to control recidivism, the extension of such FEA-aided decisions into the public sector can have severe consequences and disrupt the intent of many existing public policies and priorities.

While many public policies designed to maintain privacy and mitigate discriminatory administrative practices were passed in the 1960s and 1970s, advances in computing and the widespread storage of private data have compromised the goals of such policies. Public and private firms began to search for relationships in data to aid operational decisions with a focus on enhancing efficiency, often at the expense of other indicators of organizational well-being. For their part, public policymakers increasingly passed laws to leverage such data for policies like crime control (50-52). Concerns raised over the accessibility of such data led to new privacy laws in the 1970s and 1980s, the discussion of a “right to information privacy,” and how such rights could be protected within organizations via a “zone of privacy” for individuals and their data (53). However, privacy is a loosely defined concept that requires deeper consideration of cultural and temporal context (50), and discussions do not easily translate into law; currently we face significant concerns over information privacy and the control of data collected by private organizations and governments. With FEA-based data, such concerns only intensify.

To date, information privacy research and policy have focused on records of knowingly expressed or recorded behaviors; however, advances in AI and personal data digitalization promise to increasingly enable identification of physiological states and, by their extrapolation, internal cognitions and behavioral intent (54, 55). In short, FEA threatens to permeate a sacred boundary by imputing one’s internal thoughts or feelings based upon the often pre-cognitive or involuntary physiological responses individuals have to stimuli (56). Warren and Brandeis foreshadowed these concerns in their seminal work that established the basis for modern U.S. privacy legislation (57).

“The circumstance that a thought or emotion has been recorded in a permanent form renders its identification easier, and hence may be important from the point of view of evidence, but it has no significance as a matter of substantive right. If, then, the decisions indicate a general right to privacy for thoughts, emotions, and sensations, these should receive the same protection, whether expressed in writing, or in conduct, in conversation, in attitudes, or in facial expression.” (p. 206, emphasis added)

Much like the implementation of time-and-motion studies in the early 1900s that enabled scientific management, the use of digital databases since the 1960s that store and analyze personal information, and the “big data” revolution of the past two decades that enables both hyper-personalization and mass surveillance, FEA has the potential to influence the management of organizations and governing of society in ways that could
harm employees and the public. As evidenced by the continued use of the polygraph as a “lie detector” in employment, legal, and political domains, potentially life changing decisions are made despite low levels of confidence in a technology’s efficacy. Likewise, there will be a temptation to use FEA in a variety of big data applications that record facial behavior and then extrapolate FEA data to provide predictive analytics that impact employment decisions – or worse.

The simplified analysis offered by FEA may be seen as a strength by those who would use this tool in public, non-profit, and private sector decision making. However, the large-scale uses of such a technology would introduce and perpetuate biases if used to inform decisions without proper skepticism and regulation to consider context and other influences. For instance, a hostile off-camera interrogator will influence facial response behavior with implications for personnel decisions in organizations, and importantly, prosecution, judgment, and sentencing decisions in the criminal justice system. FEA might likewise exacerbate pre-existing biases against rural residents and lower-status individuals due to variations in facial behavior, resulting in long term societal impacts. Perhaps the greatest cause for concern with this (and other) new biometric technologies is an emerging threat to the boundary between what we make visible to an observer and the thoughts and feelings we keep to ourselves.

Patrick A. Stewart, Ph.D.
Department of Political Science
435 Old Main
University of Arkansas, Fayetteville
Fayetteville, AR 72701

Jeffrey K. Mullins, Ph.D.
Department of Information Sciences, Walton College
University of Arkansas, Fayetteville

Thomas J. Greitens, Ph.D.
Department of Political Science & Public Administration,
Central Michigan University

References