A PICTURE IS WORTH A THOUSAND WORDS: THE EVIDENTIARY PROMISE OF EMOJI-BASED SENTIMENT ANALYSIS

Taylor Lain

WHAT ARE EMOJI?

"A picture is worth a thousand words." This is an oft-used turn of phrase that, while cliché, has never in our legal history appeared as relevant as it does today. Over centuries—even millennia—images have given people outlets for cultural expression and been regarded as ways to convey a variety of intense, complex, and even hidden emotions and meanings. However, up until the past three decades,[1] one might have had to cleverly paint, draw, sculpt, or photograph such an image to convey these emotions to others. Today, digital media systems like iOS and Facebook keyboards on modern-day devices provide a way to convey them with merely the push of a button, through pictographs called *emoji*.[2]

An *emoji*, as defined by Oxford Dictionaries online, is "[a] small digital image or icon used to express an idea or emotion."[3] A more comprehensive definition provides that an emoji is "any of various small images, symbols, or icons used in text fields in electronic communication . . . to express the emotional attitude of the writer, convey information succinctly, communicate a message playfully without using words, etc."[4]

While a relatively new phenomenon, emoji are now generally considered—particularly by millennials—vital to human expression through textual forms of communication, especially through all forms of social media.[5] Emoji are used across a wide variety of communication platforms, such as Android and iOS text systems, Facebook, Instagram, Snapchat, and the like.[6] As of March 2019, there were approximately 3,019 emoji available for use in the Unicode Standard,[7] including a wide array of gender, skin tone, and

---

[1] *See* Drake Baer, *Emoticons have basically saved human communication*, BUS. INSIDER (Sept. 24, 2015, 12:50 PM), https://www.businessinsider.com/emoji-were-invented-33-years-ago-heres-why-theyre-so-crucial-today-2015-9.

[2] *See id.*

[3] *Emoji*, LEXICO, https://www.lexico.com/en/definition/emoji (last visited Aug. 28, 2019).

[4] *Emoji*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/emoji (last visited Aug. 28, 2019).

[5] *See* Baer, *supra* note 1; *see also* Gemma Church, *The hidden benefits of emojis*, CREATIVE DIGEST (Feb. 19, 2016), http://www.creativedigest.net/the-hidden-benefits-of-emoji/.

[6] *See* EMOJIPEDIA, https://emojipedia.org (last visited Aug. 28, 2019).

[7] The Unicode Standard, according to its official website, "is a character coding system designed to support the worldwide interchange, processing, and display

flag options.[8] Over five billion emoji are sent daily on Facebook Messenger alone, and by mid-2015 over half of all Instagram comments contained emoji.[9]

WHY ARE EMOJI SO PREVALENT?

Emoji confer many advantages upon their users. First, they allow users to provide emotional context for their statements.[10] The major benefit of text communication lies in that it generally does not require face-to-face interaction. Yet this leaves conversational partners communicating via text at an interpretational disadvantage because they lack access to visual and auditory cues that are normally gathered through the course of face-to-face interaction: changes in expression, tone of voice, etc.[11] Emoji offer a means by which a recipient can at least glean more obvious sentimental meaning from visual cues within the text.[12] With an emoji, a user can highlight a singular emotion or mixture of emotions that should be conveyed either by the associated text or without textual accompaniment.[13] A user can also clarify a predominant emotion or mixture of emotions that she intends to convey through a piece of text that could be ironic or have ambiguous sentiment attached.[14]

of the written texts of diverse languages and technical disciplines of the modern world," as well as "support[] classical and historical texts of many written languages." *About the Unicode® Standard*, UNICODE STANDARD, http://unicode.org/standard/standard.html (last visited Sept. 10, 2019). The system is updated regularly by a Unicode Technical Committee. *Id.*

[8] EMOJIPEDIA, https://emojipedia.org/stats/ (last visited Aug. 28, 2019).

[9] *Id.*

[10] *See* Baer, *supra* note 1.

[11] *See* Church, *supra* note 5; *see also* Alexander Hogenboom et al., *Exploiting Emoticons in Sentiment Analysis*, CONFERENCE: PROCEEDINGS OF THE 28TH ANNUAL ACM SYMPOSIUM ON APPLIED COMPUTING 1, 2–3 (2013), https://www.researchgate.net/publication/262255351_Exploiting_emoticons_in_sentiment_analysis.

[12] *See* Church, *supra* note 5; *see also* Hannah Miller et al., *Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication*, AAAI PUBLICATIONS, ELEVENTH INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA 1, 2 (2017), http://www.brenthecht.com/publications/icwsm17_emojitext.pdf.

[13] *See Emoji*, MERRIAM-WEBSTER, *supra* note 4.

[14] *See* Hogenboom et al., *supra* note 11, at 3; *see also* Benjamin Weissman & Darren Tanner, *A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension*, PLOS ONE 1, 2 (2018), http://dx.plos.org/10.1371/journal.pone.0201727; John G. Browning & Gwendolyn Seale, *More Than Words: The Evidentiary Value of Emoji*, 57 No. 10 DRI FOR DEF. 34 (2015).

Finally, a user may incorporate emoji to integrate background emotional context into what most people would consider to be a neutral factual statement.[15] Consider the examples provided in **Table 1** below.

Table 1: Expressions of Emotional Content in Texts with Emoji

Statement	Purpose of Emoji/Intended Emotion
I'm so happy that my best friend is coming home today! 😊	An emoji is added to emphasize the user's happiness associated with a positive statement.
Oh, that waiter's service was the BEST. 🙄	An emoji is added to denote sarcasm in the user's statement. While the grammatical construction of the text might be construed to convey a positive message, the original statement is intended to be negative.
I didn't get to buy concert tickets in time. 😭	An emoji is added to indicate that the user is upset about a neutral factual statement.

Second, emoji serve as an efficient means by which tangible things, such as people, places, and objects; events, such as weddings or parties; intangible ideas, like emotions or qualities; and various other ideas can be expressed concisely and efficiently, with very little explanation and with relatively profound depth of meaning.[16] As Casper Grathwohl, president of dictionaries at Oxford Dictionaries, stated, "[t]raditional alphabet language has a hard time keeping up and adapting to our needs here. The idea of a pictogram communication form like emoji, coupled with traditional alphabet languages, allows for a deeper subtlety and richness."[17]

In this respect, a picture is indeed worth a thousand words. Consider again the second example in **Table 1**. Note how it takes significantly more time and verbiage (and, arguably, would do so at the expense of confusing the reader) to explain the irony or sarcasm within the user's statement through text. The user, to clarify her meaning, would have to add to the statement something akin to the

---

[15] *See* Hogenboom et al., *supra* note 11, at 3.

[16] *See* Church, *supra* note 5.

[17] *Id.* (quoting Casper Grathwohl).

following: *NOT. That service was terrible*. Instead, the user chooses to use a single graphic to denote the sarcasm she could not convey via the original text alone; she thus saves time and effort without sacrificing meaning.

Third, and flowing from the hypothetical given above, emoji offer users a means by which myriad emotions and other information may be communicated nonverbally.[18] In fact, with the variety of emoji available now on different operating systems, it is possible for conversational partners to carry out entire conversations in emoji. Users may apply this benefit in a variety of ways, from simply sending coy notes and puzzles to friends to exchanging covert messages between criminal co-conspirators.[19] Given the variety of possible interpretations for a specific string of emoji, and the variety of meanings to which even a single emoji may correspond, the significance of a certain emoji grouping may be known only by those who are parties to the communication and thus serve as an effective means of keeping messages private.[20] Consider the peach emoji (🍑), which is used to refer to an actual peach only seven percent of the time it is used,[21] and the string of emoji and two (of arguably many) potential interpretations provided in **Figure 1** below.

Figure 1: Ambiguity in Meaning of a String of Emoji

😂 💀 🐀 [Face with Tears of Joy + Skull + Rat]

Interpretation #1. *I laughed at the dead rat.*

Interpretation #2. *I died laughing at the rat.*

Given their ample benefits, emoji will likely continue to be used frequently in future communications.

---

[18] *See* Browning & Seale, *supra* note 14.

[19] *See id.*

[20] *See* Hadley Mendelsohn, *Surprise—These Emoji Meanings Aren't What You'd Expect*, MYDOMAINE (last updated May 6, 2019), https://www.mydomaine.com/emoji-meanings.

[21] The peach emoji "is much more likely to have a sexual connotation or be used as a shorthand for butt, than it is to refer to the actual fruit." Hamdan Azhar, *How We Really Use The Peach*, EMOJIPEDIA (Dec. 16, 2016), https://blog.emojipedia.org/how-we-really-use-the-peach/.

WHAT IS SENTIMENT ANALYSIS?

Sentiment analysis, also known as *opinion mining*, is "the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from a text."[22] Sentiment analysis is the process by which artificial intelligence (AI), usually in the form of some combination of machine learning, natural language processing, and other applications, "is used to determine . . . whether a given text is subjective (presents a positive or negative sentiment) or objective and can . . . [aid in] classifying positive or negative sentiments."[23] In other words, it is a means by which AI is used to analyze documents or texts and determine (1) whether any sentimental content is contained therein, and (2) whether the expressed sentiment, if present, is positive or negative.[24]

Two methods, or some combination thereof, have generally been adopted for sentiment analysis purposes. The first, a *lexicon-based* or *semantic* approach, involves generating a dictionary of words for which polarities—the positive, negative, or neutral contents of words—are assigned numerical values.[25] During analysis, algorithmic adjustments are made to account for natural linguistic constructions.[26] Systems using the lexicon-based approach then identify the remaining words in the sentence and assign them their associated polarity values.[27] Finally, they add the polarity values and generate a score for the entire text that indicates if the polarity of the text is, as a whole, positive, negative, or neutral.[28]

The second approach, based in *machine learning*, works by training digital classifiers with texts that have been annotated with their overall polarities.[29] Each text is usually broken down into groups of words and analyzed in pieces. This eventually allows for generation and modification of algorithms that can identify semantic

---

[22] Novak et al., *Sentiment of Emojis*, PLOS ONE 1, 2 (2015), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296.

[23] Yi-Hung Liu & Yen-Liang Chen, *A two-phase sentiment analysis approach for judgment prediction*, 44 J. INFO. SCI. 594, 596 (2017), http://journals.sagepub.com/doi/pdf/10.1177/0165551517722741.

[24] *See* Hogenboom et al., *supra* note 11, at 2; *see also* Julio Villena, *An Introduction to Sentiment Analysis (Opinion Mining)*, MEANINGCLOUD (Oct. 13, 2015), https://www.meaningcloud.com/blog/an-introduction-to-sentiment-analysis-opinion-mining-in-meaningcloud.

[25] *See* Villena, *supra* note 24.

[26] *Id.*

[27] *Id.*

[28] *Id.*

[29] *Id.*

features that contribute to the overall polarity of the text.[30] By analyzing enough classified texts, these algorithms can become sophisticated enough to assign polarity values to previously unknown and unclassified texts.[31]

The two methods described above each have their own sets of benefits and problems. However, in general, either method only allows for up to a certain level accuracy in polarity assignment, even with the most finely-tuned and well-developed algorithms.[32] While not ideal, this seems logical, as humans are often unable to determine the underlying connotations of statements by simply reading text. This logic has been asserted in recent criminal cases involving social media evidence by defense attorneys arguing that much of the subtext within such conversations and posts is not conveyed by reading text alone.[33] As a result, it would be prudent for sentiment analysis to incorporate techniques that analyze emoji, through which we are better able to understand the underlying emotion in written statements. This is what researchers have recently targeted to further improve sentiment-analysis accuracy, both through lexicon-based and machine learning-based approaches.[34]

EMOJI-BASED SENTIMENT ANALYSIS: USEFUL OR NOT?

Sentiment analysis has historically been harnessed in research efforts that target the examination of documents, which range from movie critiques, to judicial opinions[35] and legal "Blawg" posts,[36] to tweets and posts on social media.[37] However, as of this decade, research efforts have been increasingly devoted to identifying the usefulness of incorporating the inherent sentiment of emoji into interpretation of textual bodies.[38] Incorporating emoji sentiment into formerly strict lexicon- or machine learning-based sentiment analysis approaches—a technique to which we will now refer as *emoji-based sentiment analysis* (EBSA)—provides many advantages:

---

[30] *Id.*

[31] *Id.*; *see also* Hogenboom et al., *supra* note 11, at 2.

[32] *See* Hogenboom et al., *supra* note 11, at 2, 7.

[33] *See* Browning & Seale, *supra* note 14.

[34] *See, e.g.,* Hogenboom et al., *supra* note 11, at 5–7.

[35] *See* Liu & Chen, *supra* note 23, at 594.

[36] *See* Jack G. Conrad & Frank Schilder, *Opinion Mining in Legal Blogs*, PROCEEDINGS OF THE 11TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 231, 231 (2007), http://www.conradweb.org/~jackg/pubs/ICAIL07_Conrad_Schilder.pdf.

[37] *See* Hogenboom et al., *supra* note 11, at 2.

[38] Hogenboom et al., *supra* note 11, at 2; Novak et al., *supra* note 22, at 2–3.

1. ***Emoji tend to be domain and topic independent.***[39] Emoji, while typically used in informal settings, generally carry the same type of polarity in meaning (or meanings) no matter the text in which they are placed.[40] This is particularly true for *emoticons*, which are emoji that are used to reflect specific emotions (such as the smiling or crying faces) as opposed to objects or places, which are often considered to be neutral in polarity. This is advantageous because it reduces the amount of topic- or domain-based grammatical and lexical tailoring that must be performed with strict machine learning- or lexicon-based approaches, as emoji connotations are generally consistent no matter the style of writing and no matter the differences in language, dialect, or vernacular.[41]

2. ***Emoji are easier to identify as the dominant emotion in a given text.***[42] As demonstrated in **Table 1**, emoji, and particularly emoticons, are often able to definitively identify the main emotion conveyed by a statement.[43] One study indicated that, while emoji are generally placed at the ends of sentences, the emotions conveyed by those emoji tend to outweigh those emotions conveyed by the language of the sentences themselves.[44] As a result, they often clarify the underlying meaning behind any given statement, particularly in informal or personal settings, and may provide information about emotions that are not apparent from the text itself.[45]

3. ***Emoji can eliminate confusion over or misunderstanding of textual meaning.***[46] Emoji, because they generally clarify the dominant emotion within a statement, have the potential to resolve misunderstandings of irony or sarcasm.[47] They clarify when polarity or tonal changes occur within a text and often identify or clarify the final takeaway meaning of multiple polarity changes within a statement or group of statements.[48]

HOW MIGHT WE APPLY EMOJI-BASED SENTIMENT ANALYSIS?

---

[39] *See* Hogenboom et al., *supra* note 11, at 2; *see* Novak et al., *supra* note 22, at 2–3.
[40] *See* Hogenboom et al., *supra* note 11, at 2.
[41] *See id.*; *see* Novak et al., *supra* note 22, at 2–3.
[42] Hogenboom et al., *supra* note 11, at 3.
[43] *Id.*
[44] *Id.*
[45] *Id.*
[46] *Id.*; *see also* Weissman & Tanner, *supra* note 14, at 2.
[47] *See* Hogenboom et al., *supra* note 11, at 3; *see also* Weissman & Tanner, *supra* note 14, at 21.
[48] *See* Hogenboom et al., *supra* note 11, at 3.

Past studies have attempted to incorporate sentiment analysis into the legal sector in a variety of ways, from mining opinions regarding legal developments discussed on Blawgs,[49] to identifying and extracting the fundamental arguments of various cases, to predicting case outcomes. [50] However, given that most of the source documents used for all of the above types of analyses are formal in nature, there is very little data contained within these documents that would require the assistance of EBSA-based methods. The most likely and useful application of such methods have been highlighted by more recent studies of informal texts, like social media posts.[51]

As stated previously, emoji tend to provide context for statements that might otherwise be unavailable to a reader. Thus emoji have recently become generally indispensable indicators in the analysis of informal textual evidence often submitted to courts in criminal cases, including drug crimes, rape and statutory rape, homicides, and even in cases pertaining to acts or threatened acts of domestic or international terrorism.[52] In the realm of civil suits, emoji have served as evidence in cases related to intellectual property, workplace discrimination, contract violations, and more.[53] However, as of now, the evidentiary value of emoji is necessarily confined to the documents in which they are contained when they are submitted to courts.[54] But what if that evidentiary value was not so confined?

The potential wealth of polarity information contained within EBSA is much better suited for use in a predictive capacity. By allowing algorithms to more accurately and rapidly identify emotionally-charged content, EBSA conceivably could be used to supplement current efforts to track online posts in real time, and to isolate and identify current or potential criminal offenders or national security threats, based on expressed emotions and opinions.[55] In particular, it could be used to identify increasing

---

[49] *See* Conrad & Schilder, *supra* note 36, at 231.

[50] *See* Liu & Chen, *supra* note 23, at 594.

[51] *See* Hogenboom et al., *supra* note 11, at 1.

[52] *See* Browning & Seale, *supra* note 14; *see also* Lucas A. Ngoge & Joseph O. Orero, *Mapping of Terrorist Activities in Kenya using Sentiment Analysis*, PAN AFRICAN CONFERENCE ON SCIENCE, COMPUTING AND TELECOMMUNICATIONS (PACT) 2017 1, 1 (2017), https://su-plus.strathmore.edu/handle/11071/5186.

[53] *See* Browning & Seale, *supra* note 14.

[54] *Id.*

[55] *See id.*; *see also* Ngoge & Orero, *supra* note 522, at 1; Kevin Sullivan, *Three American teens, recruited online, are caught trying to join the Islamic State*, WASH. POST (Dec. 8, 2014), https://www.washingtonpost.com/world/national-security/three-american-teens-recruited-online-are-caught-trying-to-join-the-islamic-state/2014/12/08/8022e6c4-7afb-11e4-84d4-7c896b90abdc_story.html?utm_term=.3dc46531d47f.

levels of distress or discontent over time. For example, suppose that the following posts about her best friend *Y*, provided in **Figure 2**, appeared on individual *X*'s Facebook feed.

Figure 2: Individual *X*'s Sample Facebook Timeline (With Emoji)

March 1, 2019: *Y* is a cool person. 😎

March 15, 2019: How could *Y* do this? 😕 😣

March 16, 2019: *Y* is going to pay. ☠️

To a reader, these posts appear to demonstrate a trend on individual *X*'s timeline indicative of growing feelings of animosity toward her friend, *Y*. Now, imagine viewing the same posts without the attached emoji, as provided in **Figure 3**.

Figure 3: Individual *X*'s Sample Facebook Timeline (Without Emoji)

March 1, 2019: *Y* is a cool person.

March 15, 2019: How could *Y* do this?

March 16, 2019: *Y* is going to pay.

While most human readers would probably recognize *X*'s developing discontent from the text of her statements alone, each of those statements or questions, when read in isolation, are actually relatively neutral in character—and likely would be analyzed as such through typical lexicon- or machine learning-based sentiment analysis algorithms.[56] There are few indicators within each statement of major polarity value—particularly, positive or negative adjectives, adverbs, nouns, verbs, etc.—and the intended emotional value of each statement or question would be hard to discern from each statement read in isolation. Incorporating, through EBSA, the emoji present in **Figure 2** would thus be of substantial value in determining the connotative meaning of each of *X*'s posts. Analyzing this content over time through EBSA algorithms would

paint a clearer picture of developing hostility between friends and, taken to an extreme, a propensity towards vengeance or, perhaps, criminal activity that *X* might be willing to conduct to make *Y* "pay."

---

[56] *See* Villena, *supra* note 24.

EBSA could also be used to identify emerging or continuing support for radical, militant political or religious groups that, in the eyes of the government, threaten national security or may be prone to acts of domestic or international terrorism.[57] For example, suppose individual *Z* posted the following on her Twitter feed, as shown in **Figure 4**.

Figure 4: Individual *Z*'s Sample Twitter Feed (With Emoji)

May 28, 2014 at 8:00 PM: About to watch Saleel Sawarim. 😁

May 28, 2014 at 9:00 PM: The end of Saleel Sawarim. 🤍 😵[58]

Each post, viewed in isolation, looks like a neutral declaration of an action or—with respect to the latter post—merely the provision of a noun. However, the emoji provided along with each textual statement provides significantly more context, indicating that *Z* is reacting positively towards—or perhaps even enjoyed and supported—Saleel Sawarim, a depiction of grotesque violence posted as propaganda by the Islamic State.[59] Thus, the connotations behind the *emoji* attached to the above Tweets would likely trigger national security enforcement alarm bells, while the *text* of the Tweets alone might not. Identifying a pattern of similar emoji-laden posts over time through EBSA methods could thus efficiently identify individuals who might later engage in acts of terrorism.

This is not a far cry from what already occurs on social media platforms. For example, Facebook is well-known to have a wide array of AI-driven algorithms that continually examine, collect, and take advantage of data from users' posts.[60] Such algorithms are also capable of analyzing and determining correlations between a user's interests and outlooks and his or her posted or algorithmically-determined race, gender, socioeconomic status, sexuality, disabilities, and other personal and private

---

[57] *See* Browning & Seale, *supra* note 14.

[58] The latter of these two posts was recreated based on a Tweet FBI officials used to identify and locate the Khan siblings, who were eventually arrested before they boarded a flight to Turkey to join the militant Islamic State in 2014. Saleel Sawarim is an hour-long Islamic State propaganda video that depicts violent acts, such as beheadings, through video and photos. *See* Sullivan, *supra* note 55.

[59] *See* Sullivan, *supra* note 55.

[60] *See* Sam Levin, *Facebook told advertisers it can identify teens feeling "insecure" and "worthless*,*"* Guardian (May 1, 2017, 3:01 PM), https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens.

characteristics.[61] Using these correlations, the algorithms make generalized assumptions regarding users' content and demographic characteristics, using one as a proxy for the other.[62] This allows these programs to then assign labels to users for purposes of marketing and recommending content.[63] As such, governments, including the U.S. government, have increasingly requested such labeling data from Facebook,[64] often with associated demands to keep their requests hidden from Facebook users.[65] And Facebook is likely not the only target, given that, as early as 2014, law enforcement agencies had already started focusing heavily on identifying national security threats through Twitter and other social media-site content.[66]

EBSA, in a more fully-developed form, could only support such sweeping tracking efforts. Because emoji tend to be less content-oriented and more widely applicable to a variety of contextual circumstances, tracking emoji rather than words offers AI algorithms—which would have to focus significantly less processing power on analyzing sentence structure and vocabulary to capture a user's meaning—a much more readily-accessible and streamlined way to hone in on possible criminal and terrorist threats based on emotionally-charged content.[67] For example, one could conceivably target users who make use of the bomb emoji (💣) or the angry face emoticon (😠) and use identifying information to track them and others with whom they interact regularly as a means of monitoring perceived threats. This would not be hard to do, given

that there are publicly-accessible programs that already track every emoji used on platforms like Twitter in real time.[68] Further, the U.S.

---

[61] *See* Kalev Leetaru, *Facebook As The Ultimate Government Surveillance Tool?*, FORBES (Jul. 20, 2018, 3:15 PM), https://www.forbes.com/sites/kalevleetaru/2018/07/20/facebook-as-the-ultimate-government-surveillance-tool/#68c56a892909.

[62] *See id.*

[63] *See id.*; *see also* Levin, *supra* note 60.

[64] *See* Jon Russell, *Government requests for Facebook user data continue to increase worldwide*, FORBES (Dec. 18, 2017, 11:03 PM), https://techcrunch.com/2017/12/18/government-requests-for-facebook-user-data-continue-to-increase-worldwide/.

[65] *See* Chris Sonderby, *Reinforcing Our Commitment to Transparency*, FORBES (Dec. 2017), https://newsroom.fb.com/news/2017/12/reinforcing-our-commitment-to-transparency/amp/.

[66] *See* Sullivan, *supra* note 55.

[67] *See* Hogenboom et al., *supra* note 11, at 2; *see also* Novak et al., *supra* note 22, at 2–3.

[68] EMOJITRACKER, http://emojitracker.com/ (last visited Apr. 18, 2019).

would simply be following in the footsteps of other nations attempting to do the same thing.[69]

SO, WHAT IS THE PROBLEM?

While incorporating EBSA into a government's repertoire seems like a promising mechanism for monitoring public safety threats, doing so presents a variety of other issues. The first set of problems specifically relate to emoji interpretation, while the second set are more general questions as to legal and ethical priorities users of EBSA will have to consider before incorporating such advanced monitoring capabilities into tracking schemes.

The interpretive value of emoji, until more research is performed, remains restricted primarily to polarity analysis of emoticons. As of yet, very little research has been performed to actually associate graphics with their meaning.[70] Even if emoji were algorithmically correlated with their meanings, the meanings of some particular emoji, like the peach emoji,[71] will be culturally dependent[72] or necessarily tied to the linguistic context in which they appear. A further, and still more problematic, extension of this idea is that the meanings of some emoji may be tied to individual interpretation—or misinterpretation[73]—as certain parties construct entirely new and individualized denotations and connotations for particular emoji in extremely specific factual scenarios.[74] As such, EBSA, even in its more advanced stages, may necessarily be subject to producing false-positives and false-negatives: some benign content mistakenly may be seen as indicative of public safety concerns, while other, more malicious content may go undetected.

Further, the prospect of incorporating EBSA into threat-tracking strategies brings a number of concerns regarding constitutional rights violations to the fore—specifically, those involving rights of privacy, equal protection, due process, and freedom of speech. Burgeoning AI developments, including sentiment analysis and EBSA applications, require immense

amounts of user data to build and test accurate models that target and track still more user data.[75] Accordingly, these kinds of

---

[69] *See* Ngoge & Orero, *supra* note 52, at 1.
[70] *C.f.* Hogenboom et al., *supra* note 11; Novak et al., *supra* note 22.
[71] *See* Azhar, *supra* note 21.
[72] *See* Mendelsohn, *supra* note 20.
[73] *See* Miller et al., *supra* note 12, at 9–10.
[74] *See* Mendelsohn, *supra* note 20.
[75] *See* Hogenboom et al., *supra* note 11, at 1.

applications tend to pull data from sources where it is quickly generated and readily accessible: social media accounts.[76]

While some of that social media user data is arguably public, issues arise when data is "deleted" (but can still be accessed through mining efforts) or when users set privacy limits to restrict viewership of their data. At these points, even if that "private" data is still somehow accessible, should it be deemed off-limits? If not, how much of that data should be considered public? Should only overt posts be considered public, or should analytical inferences made based on that data by social media corporations be considered public, too? If the latter is allowable, to what extent? Further, if that data is collected by the federal government, should the government be allowed to impose non-disclosure restrictions on corporations to prevent users from being notified that their user data has been sequestered? It is possible that such data mining requests could arguably be described as unreasonable searches and seizures, particularly given that many of the analytical inferences made with that data are not public knowledge, but rather private information.[77]

These questions also tie into related ones of discrimination against particular populations based on analytical inferences regarding their demographic characteristics and sentimental content. Setting aside the implications of numerous stereotypes incorporated into such analyses in the first place,[78] we then face the problem of government and law enforcement officials using such data to take preemptive, preventative measures to monitor populations that they (or a social media outlet) label as "at risk." These efforts necessarily become discriminatory towards particular groups of people. Discriminatory monitoring tends to beget higher instances of identifying criminal conduct (as one is bound to find something more often when they are looking for it), thereby making it even *more* likely that those same groups will be monitored, and so on in a cycle. These particular groups with opinions or comments deemed

---

[76] *See id.*

[77] Whether courts would view such data mining requests as unreasonable searches and seizures may depend on a court's willingness to analogize them to data collection processes discussed in *Carpenter v. United States*, 138 S. Ct. 2206 (2018), or *United States v. Ackies*, 918 F.3d 190 (1st Cir. 2019). This may prove to be a continuing and contentious issue in future years that will require further discussion at a later date.

[78] This leaves wide open the question as to whether the identification of certain demographic characteristics, such as socioeconomic status, through these algorithms might be pertinent to any governmental interests, and could be performed or used without becoming discriminatory in nature. This also leaves subject to debate whether data and analytical methodologies that are not overtly based on protected characteristics or discriminatory in nature still reflect individual and societal biases that could eventually lead to the development of biased threat-tracking strategies.

"unsavory" thus, in a way, essentially become guilty of crimes that have not even occurred based on highly circumstantial, generalizing evidence derived from a few posts on Facebook or Twitter.[79] Moreover, given the ample room for misinterpretation and linguistic limitations of sentiment analysis and EBSA, the risk of *accidentally* targeting these groups is exponentially increased. As a result, whether employing such mechanisms seems appropriate is a question that must be asked before we can make full use of sentiment analysis's and EBSA's predictive capacities.

Extending from the above, too, are questions of First Amendment rights. Will people really be able to speak freely on social media platforms? If content is necessarily monitored by government and law enforcement officials looking for unsavory or unorthodox material to identify potential criminal and terrorist threats, that material is *necessarily* subjected to implicit content-based, and even viewpoint, discrimination, as the government, to determine what and whom to monitor, has to decide what language is inappropriate or dangerous. These types of decisions come perilously close to free speech restrictions and outright censorship. As such, it must be asked whether law enforcement using sentiment analysis and EBSA mechanisms may potentially violate First Amendment protections of free speech, and, even more importantly, if we even care if tracking is done, if for the purpose of public safety.

CONCLUSION

The potential for sentiment analysis, and EBSA in particular, in a predictive evidentiary capacity is extraordinarily far-reaching. It would allow government and law enforcement agencies to analyze and more accurately assess users' intended emotional commentary regarding any number of topics these agencies might find pertinent, putting them in a better place to identify burgeoning, subversive threats to public safety well before they manifest themselves. However, before we can fully take advantage of these promising capabilities, we must first tackle an age-old question that still looms large in this era of modern technology: How do we best balance our individual liberties against these public safety interests?

---

[79] The categorization and monitoring of "unsavory" opinions or comments at all implicates First Amendment protections against criminalization of viewpoints. And even if a viewpoint expressed intent to commit a particular crime, a *mens rea* on its own cannot stand to uphold a criminal conviction. Alas, the problem that this generates will likely need to be addressed at another time.