HHS Public Access

Author manuscript

*Ethics Behav*. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

*Ethics Behav*. 2021 ; 31(3): 181–192. doi:10.1080/10508422.2020.1817026.

# Considerations for the Ethical Implementation of Psychological Assessment Through Social Media via Machine Learning

**Megan N. Fleming**

Department of Psychological Sciences, University of Missouri – Columbia

## Abstract

The ubiquity of social media usage has led to exciting new technologies such as machine learning. Machine learning is poised to change many fields of health, including psychology. The wealth of information provided by each social media user in combination with machine learning technologies may pave the way for automated psychological assessment and diagnosis. Assessment of individuals' social media profiles using machine learning technologies for diagnosis and screening confers many benefits (i.e., time and cost efficiency, reduced recall bias, information about an individual's emotions and functioning spanning months or years, etc.); however the implementation of these technologies will pose unique challenges to the professional ethics of psychology. Namely, psychologists must understand the impact of these assessment technologies on privacy and confidentiality, informed consent, recordkeeping, bases for assessments, and diversity and justice. This paper offers a brief review of the current applications of machine learning technologies in psychology and public health, provides an overview of potential implementations in clinical settings, and introduces ethical considerations for professional psychologists. This paper presents considerations which may aid in the extension of the current Ethical Principles of Psychologists and Code of Conduct to address these important technological advancements in the field of clinical psychology.

### Keywords

automated assessment; Ethical Principles of Psychologists; machine learning in psychology; social media; psychology

The rise of social media usage has led to exciting new developments in many fields of health, including psychology. Among these developments is the psychological assessment of individuals' social media profiles using machine learning techniques. With these new developments, professional psychologists must take several important issues into account. This paper first provides a review of the current applications of machine learning techniques in the field of clinical psychological assessment to provide context for the need for more explicit guidelines for professional psychologists who, in the future, may choose to implement them in their practice. The second half focuses on the ethical issues to be addressed prior to the implementation of psychological assessment through social media via

Correspondence concerning this article should be addressed to Megan N. Fleming, Department of Psychological Sciences, University of Missouri, 200 South 7th Street, Columbia, MO 65203. mnf8t6@mail.missouri.edu.

machine learning. It should be noted that as machine learning via social media develops, these considerations will be dynamic as technology evolves. In order to provide context for the ethical issues relevant to clinicians who wish to implement these technologies into future practice, this article provides a brief exploration of the potential technological avenues for implementation including independent applications, or apps, (similar to those used on mobile devices such as smartphones, tablets, laptops, and desktop computers) and social media-based apps. The paper then discusses ethical considerations within the framework of these potential implementations.

## Background

The rise of social media worldwide has created unprecedented access to human behavior and social interaction. Social media includes a wide variety of platforms (i.e., Facebook, Twitter, Instagram, Whatsapp, SnapChat, Reddit, YouTube, etc.) which allow individuals to write public posts, send private messages, share photos, videos, and more. Through social media, people are able to follow the activity of friends, acquaintances, companies, and public figures. Users are able to connect with persons whom they interact with in their daily lives or other users whom they have never met, but with whom they share common interests. Social media platforms can be accessed through the use of computers and mobile apps and are generally free to use.

Social media rapidly integrated into the way people communicate and connect with one another. In the ten-year window between 2005 and 2015, the percentage of people in the United States who use social media jumped from 7% to 65% (Pew Research Center, 2015). As of July 2019, there were an estimated 3.53 billion social media users worldwide (Statista, 2019). In a 2019 survey by the Pew Research Center, an estimated 72% of adults in the U.S. reported using at least one social media platform. Social media sites are especially popular among adults between the ages of 18–29 (90% reported using at least one social media site) and 30–49 (82% reported using at least one social media site). 69% of adults in the U.S. stated that they currently used Facebook, 37% reported using Instagram, and 22% reported using Twitter. An estimated 74% of Facebook users, 63% of Instagram users, and 42% of Twitter users reported visiting the social networking site once a day with an estimated 51% of Facebook users, 42% of Instagram users, and 25% of Twitter users reporting that they visit the respective social networking platform multiple times per day (Pew Research Center, 2019a). Twitter users alone generate an estimated 5,787 “tweets,” or posts with 280 characters or less, per second (Oberlo, 2019). .The prevalence and duration of social media use in combination with the volume of data produced by social media users has made it an attractive area for the use and development of new techniques to assess persons and the contents of their profiles through machine learning.

The term “machine learning” refers to the study of computer algorithms which improve automatically through experience (Mitchell, 1997). In the public health domain, researchers have utilized machine learning methods to track the presence and prevalence of influenza in geographic regions, adverse effects or unexpected beneficial effects of pharmaceutical drugs, and social patterns associated with HIV risk or HIV transmission (Signorini, Segre, & Polgreen, 2011; Nikfarjam, 2015; Young, Yu, & Wang, 2017). Another exciting application

of machine learning for the purpose of public health is the web-based social media monitoring system, Twitcident (Abel et al., 2012). Twitcident is used to filter, search, and analyze information about real-world incidents or crises reported by people within a specified geographic location on their Twitter pages to provide real-time notifications to the general public, media, and appropriate authorities so they may respond. These methods can be applied to a broad range of phenomena and may present the field of psychology with a novel approach to assessment.

## Applications in Psychology

The presence and severity of mental disorder is typically assessed through surveys and formal diagnostic interviews by clinician. To arrive at a formal diagnosis, clinicians must rely on retrospective self-reports of their clients' symptoms which can be difficult to recall (Solhan et al., 2009; Wells & Horwood, 2004). For example, to meet formal criteria for Major Depressive Disorder (MDD) through the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), individuals are asked to recall whether they experienced a depressed mood most of the day, nearly every day and fatigue or loss of energy nearly every day over the course of two weeks (American Psychiatric Association, 2013). The availability of information on social media provided in or near real-time about persons' thoughts and emotions may help to minimize issues with recall and provide a powerful resource for researchers, clinicians, and clients alike to inform clinical diagnoses. Researchers have utilized machine learning techniques derived from the field of psycholinguistics to detect patterns of behaviors and linguistic styles that map onto public health trends in mental health, to distinguish between people who have been diagnosed with a mental disorder and those without, and even to predict the onset of depressive episodes (e.g., De Choudhury et al., 2013; Jashinsky et al., 2014; Reece et al., 2017). Psychologists have indeed utilized psycholinguistics regularly to assess mental status, identify risk for suicide, and diagnose people (Rude, Gortner, & Pennebaker, 2004; Sommers-Flanagan, 2018).

A recent review by Mohr, Zhang, and Schueller (2017) reports that machine learning techniques have been applied to content obtained from platforms such as Twitter, Facebook, and Instagram to assess depression severity and suicidality. Many researchers utilize Linguistic Inquiry and Word Counts (LIWC), an analytical method that utilizes the frequency and statistical associations between words and measures of psychological features, to assess psychological characteristics of social media users. Using this method, Schwartz and colleagues (2014) found that the language content of posts modestly predicted levels of depression in a sample of 28,749 Facebook users. Furthermore, they were able to detect seasonal changes in depression by assessing the words used in users' posts (i.e., depression levels rose in the winter compared to summer) at the population level. Similar techniques have also been applied to data from Twitter users across the United States to compare psycholinguistic patterns pertaining to suicidal ideation to rates of death by suicide reported by the Centers for Disease Control (Jashinsky et al., 2014). At the population level, the researchers involved were able to find geographic patterns in social media usage that correlated highly with national rates of completed suicides.

Researchers have also developed machine learning algorithms that can detect depression in individuals through social media. De Choudhury and colleagues (2013) were able to predict the onset of a future depressive episode with 70% accuracy in a sample of Twitter users through assessment of the volume of posts in a day by an individual user, linguistic style, and use of negative words. The algorithms used in this study were also able to distinguish between Twitter users with and without depression through the assessment of times at which posts were made, use of first-person pronouns, frequency of posts, and greater disclosure about symptoms. De Choudhury and colleagues (2016) also assessed posts and comments made in semi-anonymous mental health support groups on Reddit. Reddit is a social networking site that allows users to discuss topics through conversation threads called "subreddits." Subreddits are denoted with a "r/" followed by the topic to be discussed. People who use Reddit register using a username, which may or may not include their real names. In this way, users may choose to have their information linked to their real identities or not. The researchers utilized posts and comments from subreddits of various mental health support groups (i.e., focused on eating disorders, borderline personality disorder, depression, panic disorder, social anxiety, post traumatic stress disorder, and psychosis) and support groups for people living with suicidal ideation. The models were able to classify individual users' data into the mental health support groups versus suicidal ideation support groups with 83.5% accuracy. The researchers were also able to detect individuals who shifted from participation in the online mental health support groups to suicide watch groups with 77.5% accuracy. The researchers note, however, that there may be sample biases in data gathered from the semi-anonymous platform. For example, Reddit users in mental health support groups may create an additional "throwaway" account to discuss more sensitive symptoms such as suicidality thus complicating the model's ability to predict the transition from mental health support groups to suicide support groups (De Choudhury et al., 2016).

Reece and Danforth (2017) distinguished between Instagram users with a prior diagnosis of depression and those who did not with 70% accuracy through the analysis of qualities of the photographs they posted (i.e., hues of filters applied, number of faces present as detected through facial recognition software, brightness, and color saturation). In addition, the researchers were able to demonstrate improved ability over that of unassisted general practitioners in the diagnosis of depression. In another study, researchers demonstrated an 86% success rate using a Bag of Words approach, a strategy commonly employed to assess language through machine learning techniques, to classify Twitter users with MDD and those without (Nadeem et al., 2016). Using various algorithms, researchers have shown a promising ability to distinguish between Twitter users with posttraumatic stress disorder (PTSD) and those without PTSD (Coppersmith, Harman, & Dredze, 2014). Researchers have also been able to utilize learning algorithms to detect language patterns in the contents of users' Twitter accounts prior to their official diagnosis of depression and PTSD and performed favorably in distinguishing between persons diagnosed with depression and PTSD when compared to general practitioners (Reece et al., 2017). Algorithms have also been used to detect the presence of anxiety, attention deficit hyperactivity disorder (ADHD), bipolar, borderline personality disorder, eating disorders, obsessive compulsive disorder (OCD), schizophrenia, and seasonal affective disorder (SAD) with varying success (Coppersmith et al., 2015).

The use of machine learning to diagnose people through social media is a young field with many challenges to overcome. For example, samples may be biased toward younger, technologically savvy individuals (Chancellor et al., 2019). There are also documented age-related differences in self-disclosure of on social media platforms (Settani & Marengo, 2015). Some researchers have noted that there may be differences in social media users who self-disclose their symptoms online (Coppersmith, Harman, & Dredze, 2014). Yet research on psychological assessment of persons through social media using machine learning demonstrates promise. Many researchers discuss the potential for using such algorithms as a screening and/or assessment tool to be utilized by lay people and mental health providers alike. There are numerous benefits to implementing technologies to assess individuals' mental health via their social media profiles (i.e., therapist access to screening tools that utilize extensive information spanning years gathered in, or near, real time; potential for therapists to track clients' current symptoms; time- and cost-efficient alternative to traditional interview assessments, etc.). However, to understand the ethical dilemmas that arise from their implementation, we must first discuss the technological platforms which may deliver these assessment services.

## Potential Platforms for Implementation

There are a number of ways in which machine learning techniques may be implemented to assess the presence of mental disorder through social media. It is important to note that each method of implementation will come with unique ethical considerations. One way in which these techniques may be implemented is through web-based applications supported through social media websites. Apps supported through social media websites like Facebook allow developers to gather information about users' interactions with the application while also obtaining information about the user from their profile. For example, Park and colleagues (2013) developed a web-based application through Facebook called EmotionDiary which allowed researchers to assess participants' friends lists and demographic information. The application allowed users to watch informational videos about depression and the researchers were able to examine the connection between application utilization and aspects of the users' social media profiles. Similar applications supported through Facebook's online platform could be utilized to assess attributes of clients' Facebook profiles and posts to aid in clinical diagnosis.

Another option for implementation may be the development of third party apps that would be available to download onto a desktop computer. Unlike the previous example, these applications would be separate from social media platforms. Many social media websites such as Facebook, Twitter, and Instagram allow users to download files containing the contents of their social media profiles for their own use. It is feasible that a client could be asked to download the contents of their social media profiles onto a secure server and then upload these files to the third party app for assessment. In the future, clients may also be able to provide authorization for these third party applications to obtain information from their social media accounts.

## Ethical Considerations

Implementation of machine learning technologies in a manner that upholds the ethics of the field of psychotherapy warrants discussion. The following section will detail considerations for psychologists such as the bases for assessments, privacy and confidentiality, informed consent, record keeping, and diversity and justice.

### Bases for Assessments.

Prior to implementation of psychological assessment via social media machine learning, the field must first consider whether it is ethical to consider passive screening tactics an adequate basis for assessment. According to the APA Ethics Code Standard 9.01b (Bases for Assessments), psychologists provide opinions of the psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions. Though the APA offers guidelines for best practices in conducting assessments, the code is vague as to what constitutes an examination of individuals. Assessment using these automated means may create a gray area for psychologists as some may believe it to be unethical to rely on assessments made without thorough, in-person means.

There is also the concern that an examination of the contents of an individual's social media profile may not constitute an examination of the individual who owns the profile. Social media is designed for people to share information about their lives and communicate with others. It is a highly public form of communication and as such, a psychologist must be aware that the thoughts and behaviors a client portrays through social media posts may not be directly correlated with their thoughts and behaviors offline (Emanuel et al., 2014).

Bearing these issues in mind, it may not be inherently unethical to utilize these methods to assess or diagnose individuals. Standard 9.01b also states that a psychologist may implement these methods in their own practice in light of these limitations if they take steps to clarify how these limitations may affect the reliability and validity of their opinions and appropriately limit the nature and extent of their conclusions or recommendations to their clients based on the test results (APA, 2002). As more is learned about how the public nature of social media posts influences the validity and reliability of these assessment tools, psychologists may still choose to administer them in their own practice as long as their limitations are properly addressed with their clients. An additional approach may be to only implement assessment tools which generate diagnoses from data gathered from social media profiles through less public means (i.e., a person's search history, articles and/or pages they click on, private messages, etc.) in order to address discrepancies between someone's public online presence and actual thoughts and behaviors.

### Privacy and Confidentiality.

Psychologists have an obligation to protect confidential information obtained through or stored in *any medium*, recognizing that the extent and limits of confidentiality may be established by professional relationship (APA, 2002; Standard 4.01). With this in mind, psychologists who wish to utilize machine learning techniques must be aware that they may

not be able to guarantee client confidentiality by using machine learning techniques through social media-based applications. Social media websites such as Facebook offer developers the opportunity to build applications which may be easily linked within the Facebook platform. One advantage of this modality is that it is easily accessible by Facebook users, and it allows for ease of analysis and results. However, psychologists who wish to use assessment technologies that are administered through a similar platform must be aware that applications of this nature are supported by Facebook and information about a client's results on assessment measures may be made available by Facebook to third parties whom the client has not, or would not otherwise, authorize.

One way in which disclosure to third parties could prove harmful to an individual using these services through Facebook is through targeted advertisements. Private industry utilizes targeted advertisements through social media to identify individuals who would be most likely to utilize their services. Persons may be targeted based on their demographic information such as location, age, gender, and level of education (Patel, 2012). However, advertisers are able to make use of other sources of data to more precisely target audiences based on advertisements and links users have clicked on, pages they follow, and "activities people engage in on Facebook related to things like their device usage" (Facebook for Business, 2019). Confidentiality is a major concern to those who seek therapy and wish to avoid the implications of social stigma surrounding the utilization of mental health services (Clement et al., 2015). Given the vague nature of the information Facebook uses to target advertisements to its users, psychologists must be wary that encouraging clients to authorize assessments of their profiles through apps on social media platforms may lead to an inadvertent disclosure of their mental health status via advertisements that subsequently appear on the user's newsfeed. Psychologists should also be aware that the use of applications built through a social media platform will not be Health Insurance Portability and Accountability Act of 1996 (HIPAA) compliant (APA, 2013).

Though targeted advertisements through social media websites may represent one threat to privacy and confidentiality, this may be partially mitigated through the use of applications built on independent platforms. This strategy has already been implemented in an attempt to minimize intrusions on privacy of users of an automated chatbot named Woebot that initially delivered cognitive behavioral therapy (CBT) to young adults through Facebook's messaging app (Woebot, n.d.). Due to concerns about Facebook's access to users' conversations with Woebot, Woebot developed a stand-alone application which is available for users to download separately from Facebook. Clinicians who wish to utilize automated assessments may choose to only utilize technologies administered through stand-alone apps to address privacy concerns related to targeted ads. However, clinicians must also be aware that stand-alone applications may not necessarily be HIPAA compliant and may also authorize the access of information by third parties for maintenance and delivery of services (Karcher & Presser, 2018; APA, 2013). The psychologist would maintain an ethical obligation to remain up-to-date on applications' terms of privacy and confidentiality and, under Standard 4.02C: Discussing the Limits of Confidentiality, would be obligated to inform clients of these risks to privacy and limits of confidentiality through electronic transmission (APA, 2002; Karcher & Presser, 2018).

According to APA Standard 4.04: Minimizing Intrusions on Privacy, psychologists should only include information germane to the purpose for which the communication is made. Psychologists must be aware that the amount and breadth of information the contents of a client's social media profile would provide may exceed that which is necessary to accomplish the goal of providing an assessment. For example, use of an application which requires the download of an individual's complete profile including photographs, videos, etc. may be inappropriate for an assessment tool which utilizes natural language processing techniques that only make use of textual posts to assess individuals for mental disorder.

### Informed Consent.

Psychologists have a responsibility under Standard 9.03a: Informed Consent in Assessments of the APA Ethics Code to obtain informed consent for assessments, evaluations, or diagnostic services, which includes an explanation of the involvement of third parties and limits of confidentiality. Therefore, consent to assess the contents of an individual's social media profile must be granted by the individual. The importance of a client's understanding of the relevant limits to confidentiality is echoed by Standard 4.02a: Discussing the Limits of Confidentiality which states that psychologists must discuss the relevant limits of confidentiality and the foreseeable uses of the information generated through their professional relationship with the persons with whom they work. Some ethical challenges to assessing social media profiles through automated machine-learning methods are similar to the challenges faced by mobile health (mHealth). For example, psychologists must be prepared to discuss privacy and confidentiality policies of the online assessment tools they plan to use with their clients in a manner that may be readily understood by the individual receiving services. It would also be important to discuss the potential for third parties such as developers and other app provider personnel to access client information once the client has consented to the assessment procedure. According to APA Ethics Code Standard 4.02b: Discussing the Limits of Confidentiality, psychologists have a responsibility to discuss these potential breaches of privacy and confidentiality at the outset of the relationship and thereafter as new circumstances warrant. In order to uphold this standard, psychologists must remain diligent in reviewing changes in privacy agreements of the technological services they plan to use for assessment. Psychologists may also consider discussing the potential for these accidental disclosures to take place with their clients prior to utilizing these assessment methods.

Psychologists must also be aware of the kind of information applications and platforms delivering psychological assessment will obtain. The materials available for download from a client's social media profile may include other people's replies to the client's posts, messages to the client, "likes" from other users, and the faces of persons with whom a client interacts. If the private data of other individuals would be obtained through the authorization of a client to release their social media profile, psychologists must consider whether the informed consent of other people is warranted. Though the content would most probably be excluded from psychological analysis, other persons with whom a client interacts on social media may be uncomfortable to know that their replies, messages, photographs, or other interactions with the client's social media profile may be stored or accessed by third parties.

### Record Keeping.

According to the APA Ethics Code Standard 6.01: Documentation of Professional and Scientific Work and Maintenance of Records, psychologists have the responsibility to create, and to the extent the records are under their control, maintain, disseminate, store, retain, and dispose of records and data relating to their professional and scientific work. The amount of information that may be obtained through social media for the purpose of psychological assessment may also pose issues for record keeping. Psychologists will need to determine what type of information and how much is appropriate to be maintained in records. According to Standard 4.04a: Minimizing Intrusions on Privacy, psychologists must include only information in a written report that is necessary for the purpose of that communication. In the case of the assessment of mental disorder through social media, it may be difficult to determine which tweets, messages, posts, etc. were pertinent to a client's diagnosis therefore meriting documentation. Many social media websites allow users to download their posts, "likes," comments, search history, photographs, check-ins, etc. Though the raw form of a client's social media data may be available for documentation with a client's consent, not all available information will be pertinent for documentation.

In keeping records of psychological assessment, psychologists must also take care to maintain confidentiality in the creation, storage, access, transfer, and disposal of records under their control (6.02a: Maintenance, Dissemination, and Disposal of Confidential Records of Professional and Scientific Work; APA, 2002). Assessment using social media platforms poses unique challenges to the maintenance of this standard. For example, if a psychologist were to recommend the use of a third-party application to assess an individual through their social media contents, the psychologist would be responsible for the maintenance and security of the results of the assessment. According to Standard 6.02b: Maintenance, Dissemination, and Disposal of Confidential Records of Professional and Scientific Work, a psychologist may enter client confidential information into a database or system of records available to persons whose access has not been consented to by the recipient if a psychologist employs coding or other techniques to avoid the inclusion of personal identifiers. While a psychologist may utilize services that encrypt confidential files, it must be noted that it may not be possible to remove personally identifying information from an individual's social media content. Though certain details about an individual may be changed to attempt to conceal their identity, the contents of an individual's posts (i.e., "check-ins" at local restaurants, a status update about their place of work, photographs of the client) may not be readily altered to protect the client's privacy when using a third party app.

The use of machine learning technologies to diagnose individuals through social media poses a further challenge to the present definition of "test data" within the APA Ethics Code. Standard 9.04a: Release of Test Data states that "test data" refers to raw and scaled scores, client responses to test questions or stimuli, and psychologists' notes and recordings concerning client statements and behavior during an examination. This definition does not address whether data obtained through a client's social media page should be considered test data. It will be important for the field to discuss whether the contents of a client's social media page should be included in the definition of test data so that the raw contents may be

afforded protections similar to information obtained through the methods described in the present definition.

In addition, psychologists must remain conscious of the fact that the maintenance of social media data for the purpose of record keeping may also implicate the privacy of any individual who is “tagged” in posts by the client. Inclusion of other persons’ identifying information and social media data in test data records may create a new form of collaterals in psychotherapy. The APA Ethics Code is largely silent on the topic of traditional collaterals in psychotherapy (e.g., parents who attend a psychotherapy session with their child for the purpose of advancing the therapy of the child). Standard 10.02a: Therapy Involving Couples or Families states that when psychologists agree to provide services to several persons who have a relationship, they take reasonable steps to clarify at the outset (1) which of the individuals are clients and (2) the relationship the psychologist will have with each person. The clarification under this standard also includes a discussion of the psychologist’s role and the probable uses of the services provided or the information obtained. In a traditional setting, a psychologist may establish who is the client among persons present in the room and answer any questions that the collateral may have about their rights and role in this situation (Ellis, 2012). However, in the case of individuals whose social media data is inextricably linked to that of a client’s (e.g., an acquaintance who is tagged in a post made by the client that is deemed pertinent for documentation and test data), such persons may not even be aware that their data has been maintained for the purpose of another user’s psychotherapeutic records. From a risk management perspective, this may pose unique challenges to the psychologist if records released through a court order were to contain information about illicit behaviors by the client and uninformed collaterals. Due to the potential risks to privacy for people who interact with the client on social media and who have not consented to such procedures, psychologists who plan to utilize services that use machine learning to diagnose their clients must keep abreast of the larger conversation of whether any one individual is the owner of information shared through social media platforms.

### Diversity and Justice.

Presently, many of the machine learning techniques reviewed in the present article have been developed using the English language. Though English is the dominant language in the United States, a significant portion of the population speaks non-English languages instead of, or in addition to, English on their social media accounts. According to Principle E of the APA Ethics Code, psychologists must be aware of and respect language status and must consider this as a factor when working with members of such groups. Standard 9.02b: Use of Assessments states that psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. 9.02c: Use of Assessments states that psychologists must use assessment methods that are appropriate for an individual’s language preference. Psychologists must be aware that implementation of a machine learning assessment that utilizes natural language processing techniques may be contraindicated for individuals who are multilingual or non-English speakers.

In addition, members of various backgrounds may differ in their access to technology, rendering machine learning methodologies inappropriate. Though 72% of U.S. adults reported utilizing at least one social media site in 2019, social media usage varied based on several factors (Pew Research Center, 2019a). Perhaps the most important factor influencing social media use was age. In 2019, approximately 60% of adults aged 65 and older reported not using any social media sites, making them the largest demographic of non-social media users (Pew Research Center, 2019a). Use of the internet and social media sites also varied as a function of annual income (Pew Research Center, 2019a; Pew Research Center, 2019b). 82% of people who reported an annual income of $30,000 reported internet use in 2019, a proportion that is 16% lower than adults who reported an annual income of $75,000 or more (Pew Research Center, 2019b). Income disparities carried over into reported social media usage among adults reporting an annual salary of $30,000 or less. 32% of people within this population reported that they did not use social media (Pew Research Center, 2019a). These disparities in access may have important implications for the quality of diagnoses produced by assessments using machine learning techniques for older adults and people reporting varying annual incomes. According to Principle E, psychologists must also be aware of and respect differences based on age and socioeconomic status. Psychologists may uphold this principle by taking these factors into account when considering appropriate assessment tools.

## Conclusion

The field of psychological assessment using machine learning techniques administered via social media is both complicated and challenging. The implementation of these technologies in practice would provide psychologists with access to diagnoses generated using information with presumably less recall bias than traditional interviews. In addition, such tests could provide psychologists with information provided by clients in a prospective manner spanning several years. This may prove to be invaluable in gathering information about a client's mental health history. Administering these techniques may also alleviate burdens associated with the provision of traditional interview assessments (e.g., fees associated with the use of a professional's time, client time, etc.). Though these are exciting prospects, there are several ethical issues that must be considered and addressed prior to implementation of these techniques.

The ethical concerns outlined in this paper have implications for APA Ethics Code Task Force's initiative to create new Ethics Codes which will potentially offer clearer guidance to psychologists navigating ethical issues arising from the rapidly-developing technological landscape of the profession (APA, 2020). Specifically, psychologists as a field must discuss how to best maintain documentation of assessments using social media data, ensure that clients are informed of potential uses of their social media data, and ensure that their rights to privacy and confidentiality are upheld. Psychologists must also consider whether automated assessment using machine learning technologies constitutes a thorough examination of an individual. As apps using these technologies become available to psychologists, the extent of the impact of the public nature of social media activity on how accurately clients portray themselves and their symptoms in their posts must be clarified. The extent to which persons alter how they portray themselves will have important

ramifications for upholding Standard 9.01b: Bases for Assessments. A discussion of the ethics of conducting assessments through social media using machine learning techniques on data that may belong to persons in addition to the client (i.e., a photo with the faces of two readily-identifiable individuals) without the informed consent of all parties will also be warranted. Despite the challenges and the necessity for revisions in the Ethics Code to accommodate these considerations, machine learning and social media in psychology appear poised to integrate themselves into the services psychologists provide. As such, psychologists must be proactive in their consideration of these ethical issues so that we may continue to serve the public while upholding the principles of the Ethics Code.

## Preliminary Recommendations for Providers

1. The use of machine learning in the realm of psychological assessment is still a burgeoning field. As the implications of the use of these technologies on reliability and validity become established, psychologists must only use assessments which demonstrate good reliability and validity and be prepared to communicate any presenting limitations to inferences drawn from these assessments to their clients.
2. The public nature of social media posts may greatly impact the validity of assessments using machine learning technologies. As such, prior to implementing applications built using these technologies in their own practice, psychologists should understand which aspects of the person's social media profile will be used to generate a diagnosis. Psychologists may consider only implementing applications for use in their own practice that utilize data that are of a more private nature (i.e., a person's search history, articles and/or pages they click on, private messages, etc.) as opposed to data of a more public nature (i.e., posts to friends, comments, etc.).
3. Applications offering automated assessments of social media profiles that are supported through the social media platform itself should be avoided due to privacy and confidentiality concerns. Instead, psychologists should only choose to implement independent applications that are HIPAA-compliant into their practice.
4. Psychologists should review which aspects of an individual's profile are necessary for an automated assessment using machine learning technologies to take place. Apps that do not effectively communicate what information is necessary, or which require access to information that is more than necessary, to arrive at a diagnosis should be avoided. For example, if an application analyzes the words used in posts on a client's social media profile, then applications requesting downloads of photos of the client in addition to their text-based posts should be suspect.
5. Psychologists wishing to implement these technologies in their own practice should be prepared to discuss relevant risks to privacy to their clients who provide informed consent.

6. Psychologists should regularly monitor privacy agreements of the applications they implement to ensure the continued privacy of client's data and mitigate risks to privacy in the form of accidental disclosure due to changing privacy agreements of the chosen third-party application.

7. Psychologists should only implement applications in their own practice which use coding methods to maintain the confidentiality of client information.

8. Many machine learning technologies which assess individuals based on textual posts on social media platforms have been developed using the English language. Psychologists should either refrain from using these applications as an assessment tool for clients who are multilingual or are non-English speakers.

9. Assessments which utilize machine learning technologies to evaluate clients' social media profiles to arrive at a diagnosis are inappropriate to use with non-social media users and should not be used.

10. Psychologists should determine how to manage the data of persons inextricably linked to that of a client's data (i.e., photos or textual posts in which the client has "tagged" another social media user) and whether use of these assessment tools will necessitate some form of informed consent for these persons.

## Acknowledgments

The author is supported by the National Institutes of Health Grant T32 AA013526. Special thanks to Drs. Nan Presser and Rebecca Schwartz-Mette for their invaluable support in the preparation of this manuscript.

## References

Abel F, Hauff C, Houben GJ, Stronkman R, & Tao K. (2012, 4). Twitcident: fighting fire with information from social web streams. In Proceedings of the 21st International Conference on World Wide Web (pp. 305–308). ACM.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American PsychiatricPub.

American Psychological Association (2020, n.d.). Ethics Code Task Force. Retrieved from https://www.apa.org/ethics/task-force/

American Psychological Association. (2007). Record keeping guidelines. The American Psychologist, 62, 993. [PubMed: 18085845]

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct (2002, Amended June 1, 2017). Retrieved from http://www.apa.org/ethics/code/index.aspx

American Psychological Association (2013). Guidelines for the practice of telepsychology. The American Psychologist, 68, 791–800. [PubMed: 24341643]

Campbell L, Vasquez M, Behnke S, Kinscherff R. APA ethics code commentary and case illustrations. Washington, DC: American Psychological Association; 2010.

Chancellor S, Birnbaum ML, Caine ED, Silenzio VM, & De Choudhury M. (2019, 1). A taxonomy of ethical tensions in inferring mental health states from social media. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 79–88).

Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, ... & Thornicroft G (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychological medicine, 45(1), 11–27. [PubMed: 24569086]

Coppersmith G, Dredze M, Harman C, & Hollingshead K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In Proceedings of the

2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 1–10.

Coppersmith G, Harman C, & Dredze M. (2014). Measuring post traumatic stress disorder in Twitter. In Eighth international AAAI conference on weblogs and social media.

De Choudhury M, Gamon M, Counts S, Horvitz E. (2013). Predicting depression via social media. Proc. 7th. Int. AAAI Conf. Weblogs Social Media, Boston, pp. 128–37. Palo Alto, CA: Assoc. Adv. Artif. Intell.

De Choudhury M, Kiciman E, Dredze M, Coppersmith G, & Kumar M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 2098–2110. ACM.

Ellis EM (2012). What Are the Confidentiality Rights of Collaterals in Family Therapy?. The American Journal of Family Therapy, 40(5), 369–384.

Emanuel L, Neil GJ, Bevan C, Fraser DS, Stevenage SV, Whitty MT, & Jamison-Powell S. (2014). Who am I? Representing the self offline and in different online contexts. Computers in Human Behavior, 41, 146–152.

Facebook for Business. (2019, n.d.). About detailed targeting. Retrieved from https://www.facebook.com/business/help/182371508761821?id=176276233019487

Jashinsky J, Burton SH, Hanson CL, West J, Giraud-Carrier C, Barnes MD, & Argyle T. (2014). Tracking suicide risk factors through Twitter in the US. Crisis.

Karcher N & Presser N. (2018). Ethical and legal issues addressing the use of mobile health (mHealth) as an adjunct to psychotherapy. Ethics & Behavior, 28(1), 1–22.

Kern ML, Park G, Eichstaedt JC, Schwartz HA, Sap M, Smith LK, & Ungar LH(2016). Gaining insights from social media language: Methodologies and challenges. Psychological methods, 21(4), 507. [PubMed: 27505683]

Mitchell TM (1997). Machine learning. 1997. Burr Ridge, IL: McGraw Hill, 45(37), 870–877.

Mohr DC, Zhang M, & Schueller SM (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annual review of clinical psychology, 13, 23–47.

Nadeem M, Horn M, Coppersmith G, & Sen S. (2016). Identifying depression on Twitter. arXiv preprint arXiv:1607.07384.

Nikfarjam A, Sarker A, O'Connor K, Ginn R, & Gonzalez G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. Journal of the American Medical Informatics Association, 22(3), 671–681. [PubMed: 25755127]

Oberlo (2019, July 30). 10 Twitter Statistics Every Marketer Should Know in 2019. Retrieved from https://www.oberlo.com/blog/twitter-statistics

Park S, Lee SW, Kwak J, Cha M, & Jeong B. (2013). Activities on Facebook reveal the depressive state of users. Journal of medical Internet research, 15(10), e217.

Patel N. (2012, August 10). A Deep Dive Into Facebook Advertising. Retrieved from https://neilpatel.com/blog/deep-dive-facebook-advertising/

Pew Research Center (2015, October 8). Social Media Usage: 2005–2015. Retrieved from https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/

Pew Research Center (2019, April 10). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Retrieved from https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Pew Research Center (2019, June 12). Internet/Broadband Fact Sheet. Retrieved from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Prensky M. (2001). Digital natives, digital immigrants part 1. On the horizon, 9(5), 1–6.

Rude S, Gortner EM, & Pennebaker J. (2004). Language use of depressed and depression-vulnerable college students. Cognition & Emotion, 18(8), 1121–1133.

Reece AG, & Danforth CM (2017). Instagram photos reveal predictive markers of depression. EPJ Data Science, 6(1), 15.

Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, & Langer EJ (2017). Forecasting the onset and course of mental illness with Twitter data. Scientific reports, 7(1), 13006. [PubMed: 29021528]

Settanni M, & Marengo D. (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Frontiers in psychology, 6, 1045. [PubMed: 26257692]

Schwartz HA, Eichstaedt J, Kern M, Park G, Sap M, Stillwell D, ... & Ungar L. (2014,6). Towards assessing changes in degree of depression through facebook. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 118–125.

Signorini A, Segre AM, & Polgreen PM (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PloS one, 6(5), e19467. [PubMed: 21573238]

Solhan MB, Trull TJ, Jahng S, & Wood PK (2009). Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychological assessment, 21(3), 425–436. 10.1037/a00168690 [PubMed: 19719353]

Sommers-Flanagan J. (2018). Conversations about suicide: Strategies for detecting and assessing suicide risk. Journal of Health Service Psychology, 44, 33–45.

Statista (2019, September 6). Most famous social network sites worldwide as of July 2019, ranked by number of active users (in millions). Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Wells JE, & Horwood LJ (2004). How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychological medicine, 34(6), 1001. [PubMed: 15554571]

Woebot. (n.d.). In Facebook [Business page]. Retrieved December 3, 2019, from https://www.facebook.com/pg/HiWoebot/about/?ref=page_internal

Young SD, Yu W, & Wang W. (2017). Toward automating HIV identification: machine learning for rapid identification of HIV-related social media data. Journal of acquired immune deficiency syndromes (1999), 74(Suppl 2), S128. [PubMed: 28079723]