

## Viewpoint

# Ethical Implications of the Use of Language Analysis Technologies for the Diagnosis and Prediction of Psychiatric Disorders

Alexandre Andrade Loch<sup>1,2\*</sup>, BPhil, MD, PhD; Ana Caroline Lopes-Rocha<sup>1\*</sup>, BSc; Anderson Ara<sup>3\*</sup>, PhD; João Medrado Gondim<sup>4\*</sup>, BSc; Guillermo A Cecchi<sup>5\*</sup>, MSc, PhD; Cheryl Mary Corcoran<sup>6\*</sup>, MD; Natália Bezerra Mota<sup>7,8\*</sup>, MD, PhD; Felipe C Argolo<sup>1\*</sup>, MD, PhD

<sup>1</sup>Institute of Psychiatry, University of Sao Paulo, Sao Paulo, Brazil

<sup>2</sup>Instituto Nacional de Biomarcadores em Neuropsiquiatria, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazilia, Brazil

<sup>3</sup>Departamento de Estatística, Universidade Federal do Paraná, Curitiba, Brazil

<sup>4</sup>Instituto de Computação, Universidade Federal da Bahia, Salvador, Brazil

<sup>5</sup>IBM Thomas J. Watson Research Center, Yorktown Heights, NY, United States

<sup>6</sup>Icahn School of Medicine at Mount Sinai, New York, NY, United States

<sup>7</sup>Instituto de Psiquiatria, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>8</sup>Research Department at Motrix Lab, Motrix, Rio de Janeiro, Brazil

\* all authors contributed equally

### Corresponding Author:

Alexandre Andrade Loch, BPhil, MD, PhD

Institute of Psychiatry

University of Sao Paulo

R. Dr. Ovidio Pires de Campos 785, 4 andar sala 4N60

Sao Paulo, 05403010

Brazil

Phone: 55 11996201213

Email: [alexandre.loch@usp.br](mailto:alexandre.loch@usp.br)

## Abstract

Recent developments in artificial intelligence technologies have come to a point where machine learning algorithms can infer mental status based on someone's photos and texts posted on social media. More than that, these algorithms are able to predict, with a reasonable degree of accuracy, future mental illness. They potentially represent an important advance in mental health care for preventive and early diagnosis initiatives, and for aiding professionals in the follow-up and prognosis of their patients. However, important issues call for major caution in the use of such technologies, namely, privacy and the stigma related to mental disorders. In this paper, we discuss the bioethical implications of using such technologies to diagnose and predict future mental illness, given the current scenario of swiftly growing technologies that analyze human language and the online availability of personal information given by social media. We also suggest future directions to be taken to minimize the misuse of such important technologies.

(*JMIR Ment Health* 2022;9(11):e41014) doi: [10.2196/41014](https://doi.org/10.2196/41014)

### KEYWORDS

at-risk mental state; psychosis; clinical high risk; digital phenotyping; machine learning; artificial intelligence; natural language processing

## Introduction

In 2018, football commentator and former Liverpool defender Mark Lawrenson was alerted to a facial cancerous blemish by one of his viewers [1]. General practitioner Alan Brennan emailed England's BBC and alerted Lawrenson after watching him on TV and spotting the suspicious skin lesion. Lawrenson

successfully treated the skin cancer and later would bring the doctor to the TV show to interview and thank him. In 2020, reporter Victoria Price was on air when a spectator noticed a lump on her neck [2]. The woman promptly emailed the reporter alerting that Price should have her thyroid checked—the spectator reported she also had a neck bulging in the past that was revealed to be cancer. After exams and appointments with

oncologists, Price confirmed that the lump was a thyroid cancer and underwent an effective treatment [2]. These are two examples among many others in which signs of someone's undiagnosed disease could be noticed by a third party who seized the opportunity to alert the person about it. Such attitudes are often lifesaving, as they end up in diagnosis and effective treatment. However, what if mental illness could somehow also become perceptible?

In this paper, we discuss the intrinsic privacy protection of mental illnesses and how current technologies, specifically artificial intelligence (AI), allow us to "see" mental illness and potentially bypass this protection. By "see," we mean to view by digital means. Stigma is then addressed, as it is the main issue that makes the ability to "see" mental illness have different consequences as compared to "seeing" other illnesses. Bioethical issues related to both previous items and to the use of such technologies are discussed. We then address the interpretability of AI models, an issue that may threaten bioethical principles. Lastly, we discuss problems related to the use of such technologies outside clinical and research settings.

## *The Privacy Protection of Mental Illness*

We usually think of psychiatric illness as having an intrinsic privacy protection, since we need someone to talk about their thoughts and feelings to make a mental status assessment. However, many psychiatric disorders are somewhat apparent to the trained eye—and ears—of the mental health professional and even to lay people. Appearance and behavior are the first items in the mental status examination, a road map for mental health professionals that is equivalent to the physical exam of the general practitioner [3]. Along this examination, the content and form of what someone says is also carefully assessed, as speech is the main access we have to the patient's thoughts and feelings. These are all items of what we denote in a broad definition as communication: the first is called nonverbal communication (or nonverbal language), and the other is verbal communication [4]. Therefore, the way someone behaves and what someone says, even while not being in a psychiatric interview, can sometimes provide enough data to presume the possibility of a mental disorder. This is similar to the case with jaundice, weight loss, or lumps across the body in other illnesses, for instance. However, confirmation of the diagnosis is dependent upon further examination beyond the signs shown. This confirmation, as well as the disclosure of the diagnosis, is contingent on the patient's acquiescence to be submitted to laboratory tests, imaging, and physical examination, and to reveal their feelings and experiences by further questioning about their mental status during an interview. Nevertheless, this "privacy protection" might be overcome by technology and by the quick and recent progress in AI modeling.

To begin with, the wide use of social media has made an unprecedented amount of private data publicly available. This is not a novel issue, as it has been addressed in movies and publications in diverse fields [5,6], and was recently put in the spotlight of public debate as a consequence of privacy lawsuits against the big information technology corporations [7]. While the use of such personal big data for profiting purposes has been

unveiled, its use for mental health purposes remain largely unknown. People share images, videos, and texts on their social media, showing how they behave and what they speak and think. These are the very tools used by mental health professionals to make their diagnosis. Evidently the issues displayed in social media are not the ones investigated in a mental health consultation, but they often overlap as users frequently post their intimate feelings, share their mood, and so on, online. Besides this, AI techniques have evolved to an extraordinary level, and their machine learning (ML) algorithms for verbal and nonverbal language analyses of individuals has evolved likewise [8,9]. A study published in 2020 used language and images posted to Facebook to identify signs associated with psychiatric illness [10]. A ML algorithm was fed with 3,404,959 Facebook messages and 142,390 images across 223 participants with schizophrenia spectrum disorder or mood disorders and healthy volunteers. All data prior to the first psychiatric hospitalization was uploaded to minimize the potential confounds on social media activity of medications, hospitalizations, and relapses, and receiving a formal psychiatric diagnosis. The algorithm was able to differentiate the diagnosis using Facebook activity alone over a year in advance of the first-episode hospitalization, with areas under the curve (AUCs) varying between 0.72-0.77.

Regarding this "visibility" of severe mental disorders, in more clinical/research settings, video diagnosis frameworks have also been tested, with encouraging results. Researchers have found that neuromotor precursors of schizophrenia, for instance, can be traced back to childhood [11]. Accordingly, an analysis of brief videotape footage of children eating lunch suggested that observed movement anomalies were able to discriminate among those children who later developed schizophrenia and those who did not [12]. More recently, verbal language features extracted from video and audio recordings were shown to be important early signs of psychotic illness [13-17]. These features include discourse coherence, syntactic complexity, speech content poverty, metaphorical language, and language structural connectedness [13-17]. Bedi et al [13], for instance, showed that discourse coherence was significantly correlated with subclinical psychotic symptoms. Though derived from a small sample of individuals with at-risk mental states (ARMSs) for psychosis, their model could predict with 100% accuracy progression from the risk state to schizophrenia. This finding of lower speech connectedness in at-risk individuals was also replicated by Spencer et al [17], who used speech graphs in their analyses, another technique to assess syntactic complexity and speech content poverty. These features are commonly referred to as natural language processing and, in larger samples, were used to obtain AUCs as high as 0.88 for predicting which at-risk individuals will develop a psychotic disorder in the future. For individuals at first-episode psychosis, AUCs of up to 0.92 were obtained to predict who would receive the diagnosis of schizophrenia 6 months in advance [16]. Audio features such as pauses [18] and nonverbal behavior such as gestures [19] and movement in general [20] were also seen to be discriminative between healthy and ARMSs. Besides serious mental disorders [21], AI frameworks have also been developed to detect and classify other mental disorders, as shown by numerous publications and challenges to establish an accurate

depression estimation from audio, video, and text information [22,23].

Thus, the possibility to “see” mental disorders is, per se, an innovative technology. It could increase access to mental health care and allow for prevention, early diagnosis, and treatment, as in the cases of the illnesses cited in the beginning of this text. It could also aid clinicians in diagnosing, following-up, and prognosing their patients in their daily practice. This would greatly improve the outcome and quality of life for those afflicted with a mental illness. However, one issue distinguishes mental disorders from other illnesses, making the idea of “seeing” them less desirable. This issue is stigma.

## Stigma of Mental Illness

Mental illness carries a great deal of social stigma, which most physical illnesses do not [24]. Prejudice and discrimination toward people with mental disorders is historical, existing from before the birth of psychiatry [25,26]. Unfortunately, despite the great advances seen in psychiatry in the last decades, stigma still persists in several forms [27]. Its causes are many, but as mental disorders generate behavioral changes, they often tend to be judged by the public as a moral act of the patient. If one conceives mental illness as a strictly biological disease, with genetic causes, for instance, fear and feelings of insecurity arise toward the mentally ill, as if the person would not have control of their actions [28]. If mental illness is conceived as a sole psychosocial issue, the patients are to blame for their depressive symptoms, anxiety, and so on. The situation is worse for the group of severe mental disorders. Hallucinations and delusions seen in schizophrenia symbolize a mismatch of people with the disorder with the public’s common experience, generating fear, perceived danger, and the desire for social distancing [29-31].

As such, stigma constitutes a major treatment barrier for those with mental disorders, hampering preventative initiatives and substantially worsening outcomes. This barrier appears before treatment has begun (low mental health literacy, stereotypes endorsement, and diagnosis delay and denial) and continues thereafter (self-stigma, low treatment adherence, diagnosis disclosure) [32]. Stigma also contaminates those that surround the mentally ill, including family members, friends, and health professionals. Reproducing the prejudice seen in other illnesses such as AIDS and leprosy in the past, mental disorders are often faced as if they were contagious [33]. This worsens discrimination and social isolation as even close people tend to stay away from those with the diagnosis—something called courtesy stigma [34,35]. Due to this multifaceted burden of stigma, revealing that someone may have a depressive disorder, or that someone may be at risk of having schizophrenia in the future, is different from pointing out that someone should seek a doctor because of a suspected thyroid or skin cancer [36].

To make the stigma issue worse, there is the problem of false positives [37]. There is an ongoing debate on the accuracy of one of the most studied preventive paradigms in psychiatry, namely, the ARMS for psychosis concept (or clinical high risk for psychosis) [38]. Researchers are trying to enhance the accuracy of the ARMS criteria, as studies showed that most individuals that fall into the criteria (76%) do not develop a

psychotic disorder at all [39]. In other words, the use of ARMS criteria alone generates a large number of false positives. The use of language to classify ARMS individuals who will develop a future disorder can potentially predict up to 80%-90% of cases, as seen in some studies—though with small samples. This accuracy is similar to that of a pap smear to screen for cervical cancer, eliciting a false-positive rate of around 20% [40]. However, while pap smears are routinely used as an important preventive public health strategy and concerns are directed toward improving the false-negative ratio [41], “seeing” severe mental disorders through language analysis would still be a concern. Even though language frameworks can substantially reduce the number of false positives, the great stigma related to the condition and the uncertainty regarding intervention at this phase still hinder the implementation of preventive strategies. As such, the number of prevented cases must be weighed against the number of individuals harmed by being misdiagnosed as being at risk [42].

## Bioethical Issues

Given that mental disorders might be now “visible” to AI algorithms that analyze communication, and that there is a stigma related to these disorders, one must consider the bioethical implications. The four main principles of bioethics are (1) beneficence, (2) nonmaleficence, (3) autonomy, and (4) justice [43].

Beneficence is the principle that guides physicians to act for the benefit of patients. It also implies several other actions beyond the usual patient-doctor relationship in a clinic, such as rescuing persons in danger, removing conditions that will cause harm, and helping individuals with a disability. Beneficence is a positive concept in the sense that one has to be active, to propose actions and intervene. Nonmaleficence, on the other hand, concerns the obligation physicians have to not harm their patients, not worsen their health, not incapacitate, not cause suffering, and not deprive others of the goods of life. In practice this implicates the weighting of benefits against burdens of all interventions and treatments and in considering not to act [43]. Autonomy asserts that every person has the power to make rational decisions and moral choices, and everybody should be allowed to exercise their capacity for self-determination. The principle of autonomy branches out into three other important principles [43]:

1. Informed consent: patients must receive full disclosure and comprehend the disclosure to voluntarily agree to a medical procedure or research.
2. Truth-telling: a vital component in a physician-patient relationship, as autonomous patients have the right to know their diagnosis but also the option to forgo this disclosure.
3. Confidentiality: physicians are obliged not to disclose any confidential information given by patients to another party without their consent.

Justice is the fourth ethical principle. It encompasses the need for a fair, equitable and appropriate treatment of persons. This principle may encompass microsettings, such as adequately treating individuals in an emergency service, as well as macrosettings involving health care policies.

That being said, the use of language detection algorithms for mental health purposes may pose some bioethical dilemmas, and the use of such tools must still be approached with caution. The role of medical advice and diagnosis in one's disease trajectory is relevant, and the early detection of mental illness can enable health care practitioners to intervene and avoid negative outcomes. On the other hand, the attribution of labels can also increase the chances of self-stigma [42]. Social stigma is a great burden for people with mental disorders and is especially associated with the psychiatric diagnosis (ie, labels) [29], generating poorer outcomes and other negative consequences [44]. This is especially important considering the high rate of false positives among the previously cited ARMS condition [39]. That is, individuals (wrongly) classified as ARMS but who would never develop a mental disorder. Using the ARMS paradigm indiscriminately to diagnose individuals without properly demystifying this information and destigmatizing mental illnesses would constitute a threat to the nonmaleficence principle [42]. Besides, it would also be potentially paternalistic, harming autonomy.

To understand the biological pathways toward psychosis and to develop new treatments, research efforts are being directed to the enhancement of the predictive power of the ARMS concept [45]. This is being done by investigating biological markers or by using ML algorithms. In this sense, the use of AI for natural language processing has produced encouraging results, with a diagnostic accuracy similar or superior to other classifiers used in medicine [14,15]. However, if stigma is not addressed in a comprehensive way, no matter how few false positives there are, they will still be a concern regarding nonmaleficence. Moreover, the use of such algorithms to interpret language data also poses a threat to nonmaleficence and to autonomy, more specifically concerning the lack of comprehension of certain information given by such algorithms, which we depict below.

### Interpretability and Validity of Algorithms

The issue that some ML models are impossible to interpret has recently gained a growing interest [46]. There's an ongoing discussion on the repercussions of such algorithms for high-stakes decisions. Such models are called black box models, for their operation with the inputted variables is not completely observable. They are known to learn from subtle metadata, and this may carry the risk of hidden bias (eg, the Clever Hans problem) [47].

Exemplifying this, in 2018, a study aimed to investigate the generalizability across sites of a deep learning model to detect pneumonia in chest radiographs [48]. At first, the model performed very well in distinguishing high-risk patients from non-high-risk patients. However, upon external validation, the performance plummeted. The AI model was not learning from clinically relevant information from the images. Instead, it was learning and basing its decisions on hardware-related metadata tied to a specific x-ray machine. This machine was being used in an intensive care unit (ICU) to image mostly high-risk individuals [49]. That is, the algorithm would attribute a high-risk classification to most images coming from that ICU's

x-ray, instead of using clinical data from the x-rays themselves to make decisions. Several scholars have discussed explainability as a major problem for the use of AI in health care [50].

For some computational problems, it is hard to associate meaningful parameters with individual variables. For instance, in images examined by computer vision, each pixel is meaningless without context, while the full set of pixels taken together contain local (eg, pixels that together form a smile) and global (eg, sources of light inferred from shadow directions) characteristics. Complex models make use of several heuristics to capture abstract notions according to each application. Concepts such as objects in pictures and seasonality in time series are encoded and distributed across different structures within the model. Therefore, simple descriptions such as "anxiety increases as stress increases" are rarely possible, contrary to what happens in familiar regression methods. Since multiple conditional dependencies preclude direct statements about results, additional analytical and experimental steps are required for the interpretation of complex models [51].

In short, it is not enough only to enhance predictive power and avoid false positives but also to understand the real-world underpinnings of black box algorithms [49]. Both machine statements and human statements are congruent with a given ground truth [50]. Taking the above example, we have two statements, accordingly. First, a specific characteristic of some given x-rays is associated with a higher risk for pneumonia (machine statement). Second, we should prioritize patients with those x-rays, as they are at higher odds of having pneumonia according to the ML model (human statement). Both statements are equally used for decision-making. Nevertheless, human models are often based on causality as an aim for understanding underlying mechanisms and for intervention. While the correlation is accepted as a basis for decisions, it is viewed as an intermediate step. For instance, *why* are those specific patients at higher risk of having pneumonia? We should investigate their characteristics to understand the higher risk of pneumonia and to develop a specific antibiotic. On the other hand, ML algorithms are typically based on probabilistic models and provide only a crude basis for further establishing causal relationships. Upon opening the black box, the relationship between that specific set of x-rays and pneumonia was due to a given x-ray machine located in an ICU service that was working on many more cases of pneumonia than the other machines. That is, ML models offer important decision-making tools, namely, prioritizing those individuals. However, further investigations beyond the simple association should be conducted, opening the black box and addressing physiopathological explainability and causability.

Decisions in health care imply liability, including legal and financial repercussions. Therefore, each decision must be logically reasoned with explainable evidence [49]. AI models might be insightful for scientists, but they should also be sufficiently clear and explainable for end users to support their decisions [52]. Otherwise, it could constitute a threat to the patient's autonomy. Accordingly, traditional algorithms must handle sources of information in an interpretable manner, such as the GRACE score for acute coronary syndrome and the

Sequential Organ Failure Assessment score for organ failure in ICUs [53,54]. They map higher probabilities of a bad outcome to signs of severity (eg, abnormal values in biological markers). Clinical support decision systems based on opaque (“black-box”) algorithms must, as such, provide a clear rationale to be useful for practitioners. Besides carrying hidden bias, the use of opaque algorithms leads to a defensive medical practice. When no underlying rationale is presented [55], physicians tend to agree with the machine to avoid liability. On the other hand, interpretable outputs will help practitioners to treat their patients with fewer overlooked findings and misled predictions.

Post hoc techniques of analysis (local interpretable model-agnostic explanations, Shapley Additive Explanations, multilingual unsupervised and supervised embeddings, etc) are an option when model parameters are numerous and computational processes go beyond elementary functions and operators [51]. Specifically regarding language data, researchers should treat findings from computer based evaluations as they do with traditional indicators. It is crucial, for instance, to have representative data as a basis for normative curves for each proposed behavioral marker. How does it develop through ages; how does it change according to gender or ethnicity; what are the effects of social factors such as socioeconomic status, educational level, neighborhood, or exposure to urban violence? Before jumping to the conclusion on the relationship of some behavioral marker to a pathological factor, we first need to map and understand normative variability across cultures, languages, and countries [56,57]. For example, language structural connectedness that diminishes according to negative symptom severity under psychosis [16] also increases during typical development, being tightly associated with educational levels. Years of formal education are more important to explain this developmental trajectory than years of age to the point that illiterate adults narrate their stories with a structural pattern similar to preschool children [58].

After gaining insights on potential pathological markers and mapping on pathological confounding factors, we still need to discuss potential public policies that protect the individual rights to not be evaluated or even judged without consent.

## Legislation and Data Privacy

Otto Hahn won the Nobel Prize for discovering nuclear fission in 1939 and allowing nuclear reactors to produce enormous amounts of energy [59]. However, he is a controversial historical figure once his discovery also allowed the building of the World War II atomic bombs. Arthur W Galston studied the effects of 2,3,5-triiodobenzoic acid on the growth and flowering of soybeans. Later the military developed it into Agent Orange and used it in the Vietnam War as a chemical weapon. This led Galston to become a bioethicist and give talks on the misuse of science [60]. Likewise, internet, smartphones, social media, and search mechanisms revolutionized our relationship with knowledge and with each other as humans. However, unethical misuse of big data to control one’s exposure to information, to stimulate consumerism, and to capture someone’s time for profiting purposes are the proxies of such inventions’ perversion [61]. The number of discoveries—either scientific or not—that

got misused by third parties other than their inventor is countless. It is easy to figure out how technologies that address mental status through language can follow the same way.

Automated analysis of free speech, for instance, can establish thought disorder indexes based on what someone says or writes [13]. Additionally, these indexes can predict future serious mental disorders like schizophrenia. That is, the data fed into the analysis can be extracted from written text from books, transcripts, or other data sets that are available to the public (eg, social media or personal blogs). This raises the possibility of malicious use, given the online availability of people’s written data on the internet. Another example of the possible use of available information to infer the mental status of individuals is the Facebook study mentioned at the beginning. The algorithm could predict a future and severe mental disorder with a 72%-77% level of certainty. All these works importantly advance science and provide the perspective of useful tools to be used by clinicians and policy makers. Additionally these findings are developed in environments strictly guided by ethical standards given by ethics committees and supervised by the scientific community. However, the problem is not related to these regulated settings but when the invention goes beyond scientists.

Likewise, ethical boards review and regulate scientific studies and health professionals’ practice; apart from these settings, the law exists to oversee and penalize irregular use of big data [62]. As such, there is now a growing concern about digital privacy, especially after the awareness raised by the lawsuits against big tech claiming too much power over people’s lives and personal information a few years ago. This has led to an increase in legislation to regulate access and use of personal information, especially that which is somewhat publicly available in online social networks. However, there is always the risk of a legal gap as cutting-edge inventions are temporally ahead of legislation protecting them from misuse. This is especially evident today as new ML algorithms and technologies arise with increasing frequency. This can potentially foster discrimination of individuals with mental disorders in countries where such a gap is not covered by personal information privacy protection legislation. For instance, allowing the misuse of such technologies in job interviews, academic interviews, and so on, to dismiss people from the selection process based on preconceived ideas about mental disorders.

## Conclusion

Summarizing, new technologies derived from AI have the potential to “see” mental disorders by someone’s behavior and discourse. These technologies per se would greatly help in early detection and disease outcome. However, the historical and enduring stigma attached to mental disorders hampers the use of such tools. Fighting prejudice and discrimination related to mental disorders should constitute future directions so that stigma does not constitute a barrier for the use of these innovative technologies. Moreover, to comply with nonmaleficence and avoid the stigma, these technologies also need to have low rates of false positives in predicting someone’s possibility of future mental illness. There is a further risk that

these ML algorithms turn into black box models. This hidden bias problem could potentially harm the patients' autonomy and disclosure. So, it is necessary to clearly describe the algorithm, to use post hoc interpretation methods, and to conduct bias-checking procedures. Additionally, because of stigma and due to the high online availability of personal information on an individual's verbal and nonverbal language, information derived from the algorithms carries the risk of being misused,

such as to discriminate against individuals because of their mental health status. In this sense, awareness should be raised in regulating the use of these technologies in real-world settings. There is a challenge for legislators to catch-up with the ever-renovating new technologies and algorithms designed to decipher human behavior to prevent these inventions from being misused.

## Conflicts of Interest

NBM works at the Motrix, an EduTech startup, and has been a consultant to Boehringer Ingelheim. The other authors have no conflicts of interest to declare.

## References

1. Lawrenson M. How Football Focus viewer alerted me to cancer scare. BBC. 2018 Sep 07. URL: <https://www.bbc.com/sport/football/45451518> [accessed 2022-05-25]
2. Elassar A. TV reporter discovers she has cancer after a viewer noticed a suspicious lump on her neck. CNN. 2020. URL: <https://edition.cnn.com/2020/07/24/us/florida-reporter-cancer-viewer-lump-trnd/index.html> [accessed 2022-05-25]
3. Norris D, Clark MS, Shipley S. The Mental Status Examination. *Am Fam Physician* 2016 Oct 15;94(8):635-641 [FREE Full text] [Medline: 27929229]
4. Bavelas JB, Chovil N. Nonverbal and verbal communication: hand gestures and facial displays as part of language use in face-to-face dialogue. In: Manusov V, Patterson ML, editors. *The SAGE Handbook of Nonverbal Communication*. Thousand Oaks, CA: SAGE Publications Inc; 2006:97-115.
5. Cohen IG, Mello MM. Big data, big tech, and protecting patient privacy. *JAMA* 2019 Sep 24;322(12):1141-1142. [doi: 10.1001/jama.2019.11365] [Medline: 31397838]
6. Houser KA, Voss WG. GDPR: the end of Google and Facebook or a new paradigm in data privacy? *Richmond J Law Technol* 2018;1:25 [FREE Full text]
7. Roberts JJ. Big tech vs. big privacy lawsuits. *Fortune*. 2019. URL: <https://fortune.com/2019/02/23/big-tech-vs-big-privacy-lawsuits/> [accessed 2022-05-25]
8. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
9. Teferra BG, Borwein S, DeSouza DD, Simpson W, Rheault L, Rose J. Acoustic and linguistic features of impromptu speech and their association with anxiety: validation study. *JMIR Ment Health* 2022 Jul 08;9(7):e36828 [FREE Full text] [doi: 10.2196/36828] [Medline: 35802401]
10. Birnbaum ML, Norel R, Van Meter A, Ali AF, Arenare E, Eyigoz E, et al. Identifying signals associated with psychiatric illness utilizing language and images posted to Facebook. *NPJ Schizophr* 2020 Dec 03;6(1):38. [doi: 10.1038/s41537-020-00125-0] [Medline: 33273468]
11. Walker EF. Neurodevelopmental precursors of schizophrenia. In: *The Neuropsychology of Schizophrenia*. London: Taylor & Francis Group; 1994:119-129.
12. Schiffman J, Walker E, Ekstrom M, Schulsinger F, Sorensen H, Mednick S. Childhood videotaped social and neuromotor precursors of schizophrenia: a prospective investigation. *Am J Psychiatry* 2004 Nov;161(11):2021-2027. [doi: 10.1176/appi.ajp.161.11.2021] [Medline: 15514402]
13. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr* 2015;1:15030. [doi: 10.1038/npjpsych.2015.30] [Medline: 27336038]
14. Corcoran CM, Cecchi GA. Using language processing and speech analysis for the identification of psychosis and other disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020 Aug;5(8):770-779 [FREE Full text] [doi: 10.1016/j.bpsc.2020.06.004] [Medline: 32771179]
15. Corcoran CM, Mittal VA, Bearden CE, E Gur R, Hiczenko K, Bilgrami Z, et al. Language as a biomarker for psychosis: a natural language processing approach. *Schizophr Res* 2020 Dec;226:158-166 [FREE Full text] [doi: 10.1016/j.schres.2020.04.032] [Medline: 32499162]
16. Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. *NPJ Schizophr* 2017;3:18. [doi: 10.1038/s41537-017-0019-3] [Medline: 28560264]
17. Spencer TJ, Thompson B, Oliver D, Diederer K, Demjaha A, Weinstein S, et al. Lower speech connectedness linked to incidence of psychosis in people at clinical high risk. *Schizophr Res* 2021 Feb;228:493-501. [doi: 10.1016/j.schres.2020.09.002] [Medline: 32951966]

18. Stanislawski ER, Bilgrami ZR, Sarac C, Garg S, Heisig S, Cecchi GA, et al. Negative symptoms and speech pauses in youths at clinical high risk for psychosis. *NPJ Schizophr* 2021 Jan 22;7(1):3. [doi: [10.1038/s41537-020-00132-1](https://doi.org/10.1038/s41537-020-00132-1)] [Medline: [33483485](https://pubmed.ncbi.nlm.nih.gov/33483485/)]
19. Osborne KJ, Bernard JA, Gupta T, Dean DJ, Millman Z, Vargas T, et al. Beat gestures and postural control in youth at ultrahigh risk for psychosis. *Schizophr Res* 2017 Jul;185:197-199 [FREE Full text] [doi: [10.1016/j.schres.2016.11.028](https://doi.org/10.1016/j.schres.2016.11.028)] [Medline: [27914727](https://pubmed.ncbi.nlm.nih.gov/27914727/)]
20. Lopes-Rocha AC, Corcoran CM, Andrade JC, Peroni L, Haddad NM, Hortêncio L, et al. Motion energy analysis during speech tasks in medication-naïve individuals with at-risk mental states for psychosis. *Schizophrenia (Heidelb)* 2022 Sep 16;8(1):73. [doi: [10.1038/s41537-022-00283-3](https://doi.org/10.1038/s41537-022-00283-3)] [Medline: [36114187](https://pubmed.ncbi.nlm.nih.gov/36114187/)]
21. Harvey D, Lobban F, Rayson P, Warner A, Jones S. Natural language processing methods and bipolar disorder: scoping review. *JMIR Ment Health* 2022 Apr 22;9(4):e35928 [FREE Full text] [doi: [10.2196/35928](https://doi.org/10.2196/35928)] [Medline: [35451984](https://pubmed.ncbi.nlm.nih.gov/35451984/)]
22. Yang L, Sahli H, Xia X, Pei E, Oveneke MC, Jiang D. Hybrid depression classification and estimation from audio video and text information. In: *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. 2017 Oct Presented at: AVEC '17; October 23, 2017; Mountain View, CA p. 45-51. [doi: [10.1145/3133944.3133950](https://doi.org/10.1145/3133944.3133950)]
23. He L, Guo C, Tiwari P, Su R, Pandey H, Dang W. DepNet: An automated industrial intelligent system using deep learning for video - based depression analysis. *Int J Intelligent Sys* 2021 Oct 06;37(7):3815-3835. [doi: [10.1002/int.22704](https://doi.org/10.1002/int.22704)]
24. Kowalski R, Peipert A. Public- and self-stigma attached to physical versus psychological disabilities. *Stigma Health* 2019 May;4(2):136-142. [doi: [10.1037/sah0000123](https://doi.org/10.1037/sah0000123)]
25. Shorter E. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. Oxford, England: John Wiley & Sons; 1997.
26. Foucault M, Howard R. *Madness and Civilization: A History of Insanity in the Age of Reason*. New York: Vintage; 1988.
27. Fox AB, Earnshaw VA, Taverna EC, Vogt D. Conceptualizing and measuring mental illness stigma: the mental illness stigma framework and critical review of measures. *Stigma Health* 2018 Nov;3(4):348-376 [FREE Full text] [doi: [10.1037/sah0000104](https://doi.org/10.1037/sah0000104)] [Medline: [30505939](https://pubmed.ncbi.nlm.nih.gov/30505939/)]
28. Larkings JS, Brown PM. Do biogenetic causal beliefs reduce mental illness stigma in people with mental illness and in mental health professionals? A systematic review. *Int J Ment Health Nurs* 2018 Jun;27(3):928-941. [doi: [10.1111/inm.12390](https://doi.org/10.1111/inm.12390)] [Medline: [28942615](https://pubmed.ncbi.nlm.nih.gov/28942615/)]
29. Loch AA, Hengartner MP, Guarniero FB, Lawson FL, Wang Y, Gattaz WF, et al. The more information, the more negative stigma towards schizophrenia: Brazilian general population and psychiatrists compared. *Psychiatry Res* 2013 Feb 28;205(3):185-191. [doi: [10.1016/j.psychres.2012.11.023](https://doi.org/10.1016/j.psychres.2012.11.023)] [Medline: [23266022](https://pubmed.ncbi.nlm.nih.gov/23266022/)]
30. Woodberry KA, Powers KS, Bryant C, Downing D, Verdi MB, Elacqua KM, et al. Emotional and stigma-related experiences relative to being told one is at risk for psychosis. *Schizophr Res* 2021 Dec;238:44-51. [doi: [10.1016/j.schres.2021.09.017](https://doi.org/10.1016/j.schres.2021.09.017)] [Medline: [34598105](https://pubmed.ncbi.nlm.nih.gov/34598105/)]
31. Larsen EM, Herrera S, Bilgrami ZR, Shaik RB, Crump F, Sarac C, et al. Self-stigma related feelings of shame and facial fear recognition in individuals at clinical high risk for psychosis: A brief report. *Schizophr Res* 2019 Jun;208:483-485 [FREE Full text] [doi: [10.1016/j.schres.2019.01.027](https://doi.org/10.1016/j.schres.2019.01.027)] [Medline: [30700396](https://pubmed.ncbi.nlm.nih.gov/30700396/)]
32. Setti VPC, Loch AA, Modelli A, de Almeida Rocca CC, Hungerbuehler I, van de Bilt MT, et al. Disclosing the diagnosis of schizophrenia: a pilot study of the 'Coming Out Proud' intervention. *Int J Soc Psychiatry* 2019 May;65(3):244-251. [doi: [10.1177/0020764019840057](https://doi.org/10.1177/0020764019840057)] [Medline: [30915887](https://pubmed.ncbi.nlm.nih.gov/30915887/)]
33. van der Sanden RLM, Bos A, Stutterheim S, Pryor J, Kok G. Experiences of stigma by association among family members of people with mental illness. *Rehabil Psychol* 2013 Feb;58(1):73-80. [doi: [10.1037/a0031752](https://doi.org/10.1037/a0031752)] [Medline: [23438002](https://pubmed.ncbi.nlm.nih.gov/23438002/)]
34. Angermeyer MC, Schulze B, Dietrich S. Courtesy stigma--a focus group study of relatives of schizophrenia patients. *Soc Psychiatry Psychiatr Epidemiol* 2003 Oct;38(10):593-602. [doi: [10.1007/s00127-003-0680-x](https://doi.org/10.1007/s00127-003-0680-x)] [Medline: [14564387](https://pubmed.ncbi.nlm.nih.gov/14564387/)]
35. Wong C, Davidson L, Anglin D, Link B, Gerson R, Malaspina D, et al. Stigma in families of individuals in early stages of psychotic illness: family stigma and early psychosis. *Early Interv Psychiatry* 2009 May;3(2):108-115 [FREE Full text] [doi: [10.1111/j.1751-7893.2009.00116.x](https://doi.org/10.1111/j.1751-7893.2009.00116.x)] [Medline: [19777087](https://pubmed.ncbi.nlm.nih.gov/19777087/)]
36. Yang LH, Woodberry KA, Link BG, Corcoran CM, Bryant C, Shapiro DI, et al. Impact of "psychosis risk" identification: examining predictors of how youth view themselves. *Schizophr Res* 2019 Jun;208:300-307 [FREE Full text] [doi: [10.1016/j.schres.2019.01.037](https://doi.org/10.1016/j.schres.2019.01.037)] [Medline: [30792136](https://pubmed.ncbi.nlm.nih.gov/30792136/)]
37. Corcoran C, Malaspina D, Hercher L. Prodromal interventions for schizophrenia vulnerability: the risks of being "at risk". *Schizophr Res* 2005 Mar 01;73(2-3):173-184 [FREE Full text] [doi: [10.1016/j.schres.2004.05.021](https://doi.org/10.1016/j.schres.2004.05.021)] [Medline: [15653260](https://pubmed.ncbi.nlm.nih.gov/15653260/)]
38. Koutsouleris N, Worthington M, Dwyer DB, Kambeitz-Illankovic L, Sanfelici R, Fusar-Poli P, et al. Toward generalizable and transdiagnostic tools for psychosis prediction: an independent validation and improvement of the NAPLS-2 risk calculator in the multisite PRONIA cohort. *Biol Psychiatry* 2021 Nov 01;90(9):632-642. [doi: [10.1016/j.biopsych.2021.06.023](https://doi.org/10.1016/j.biopsych.2021.06.023)] [Medline: [34482951](https://pubmed.ncbi.nlm.nih.gov/34482951/)]
39. Simon AE, Velthorst E, Nieman DH, Linszen D, Umbricht D, de Haan L. Ultra high-risk state for psychosis and non-transition: a systematic review. *Schizophr Res* 2011 Oct;132(1):8-17. [doi: [10.1016/j.schres.2011.07.002](https://doi.org/10.1016/j.schres.2011.07.002)] [Medline: [21784618](https://pubmed.ncbi.nlm.nih.gov/21784618/)]

40. Jantzen J, Norup J, Dounias G, Bjerregaard B. Pap-smear benchmark data for pattern classification. 2005 Jan 01 Presented at: Proc NiSIS 2005: Nature inspired Smart Information Systems, EU co-ordination action; 2005; Albufeira, Portugal p. 1-9.
41. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. *Biomed Eng Online* 2019 Feb 12;18(1):16 [FREE Full text] [doi: [10.1186/s12938-019-0634-5](https://doi.org/10.1186/s12938-019-0634-5)] [Medline: [30755214](https://pubmed.ncbi.nlm.nih.gov/30755214/)]
42. Colizzi M, Ruggeri M, Lasalvia A. Should we be concerned about stigma and discrimination in people at risk for psychosis? A systematic review. *Psychol Med* 2020 Apr;50(5):705-726. [doi: [10.1017/S0033291720000148](https://doi.org/10.1017/S0033291720000148)] [Medline: [32063250](https://pubmed.ncbi.nlm.nih.gov/32063250/)]
43. Varkey B. Principles of clinical ethics and their application to practice. *Med Princ Pract* 2021;30(1):17-28 [FREE Full text] [doi: [10.1159/000509119](https://doi.org/10.1159/000509119)] [Medline: [32498071](https://pubmed.ncbi.nlm.nih.gov/32498071/)]
44. Wright A, Jorm AF, Mackinnon AJ. Labeling of mental disorders and stigma in young people. *Soc Sci Med* 2011 Aug;73(4):498-506. [doi: [10.1016/j.socscimed.2011.06.015](https://doi.org/10.1016/j.socscimed.2011.06.015)] [Medline: [21794967](https://pubmed.ncbi.nlm.nih.gov/21794967/)]
45. Montemagni C, Bellino S, Bracale N, Bozzatello P, Rocca P. Models predicting psychosis in patients with high clinical risk: a systematic review. *Front Psychiatry* 2020;11:223. [doi: [10.3389/fpsyt.2020.00223](https://doi.org/10.3389/fpsyt.2020.00223)] [Medline: [32265763](https://pubmed.ncbi.nlm.nih.gov/32265763/)]
46. Wongvibulsin S, Wu KC, Zeger SL. Improving clinical translation of machine learning approaches through clinician-tailored visual displays of black box algorithms: development and validation. *JMIR Med Inform* 2020 Jun 09;8(6):e15791 [FREE Full text] [doi: [10.2196/15791](https://doi.org/10.2196/15791)] [Medline: [32515746](https://pubmed.ncbi.nlm.nih.gov/32515746/)]
47. Lapuschkin S, Wäldchen S, Binder A, Montavon G, Samek W, Müller KR. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat Commun* 2019 Mar 11;10(1):1096. [doi: [10.1038/s41467-019-08987-4](https://doi.org/10.1038/s41467-019-08987-4)] [Medline: [30858366](https://pubmed.ncbi.nlm.nih.gov/30858366/)]
48. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med* 2018 Nov;15(11):e1002683 [FREE Full text] [doi: [10.1371/journal.pmed.1002683](https://doi.org/10.1371/journal.pmed.1002683)] [Medline: [30399157](https://pubmed.ncbi.nlm.nih.gov/30399157/)]
49. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020 Nov 30;20(1):310 [FREE Full text] [doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6)] [Medline: [33256715](https://pubmed.ncbi.nlm.nih.gov/33256715/)]
50. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;9(4):e1312 [FREE Full text] [doi: [10.1002/widm.1312](https://doi.org/10.1002/widm.1312)] [Medline: [32089788](https://pubmed.ncbi.nlm.nih.gov/32089788/)]
51. Abdullah TAA, Zahid MSM, Ali W. A review of interpretable ML in healthcare: taxonomy, applications, challenges, and future directions. *Symmetry* 2021 Dec 17;13(12):2439. [doi: [10.3390/sym13122439](https://doi.org/10.3390/sym13122439)]
52. Arbelaez Ossa L, Starke G, Lorenzini G, Vogt JE, Shaw DM, Elger BS. Re-focusing explainability in medicine. *Digit Health* 2022;8:20552076221074488 [FREE Full text] [doi: [10.1177/20552076221074488](https://doi.org/10.1177/20552076221074488)] [Medline: [35173981](https://pubmed.ncbi.nlm.nih.gov/35173981/)]
53. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care* 2019 Nov 27;23(1):374 [FREE Full text] [doi: [10.1186/s13054-019-2663-7](https://doi.org/10.1186/s13054-019-2663-7)] [Medline: [31775846](https://pubmed.ncbi.nlm.nih.gov/31775846/)]
54. Tang EW, Wong C, Herbison P. Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome. *Am Heart J* 2007 Jan;153(1):29-35. [doi: [10.1016/j.ahj.2006.10.004](https://doi.org/10.1016/j.ahj.2006.10.004)] [Medline: [17174633](https://pubmed.ncbi.nlm.nih.gov/17174633/)]
55. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics* 2020 Mar;46(3):205-211 [FREE Full text] [doi: [10.1136/medethics-2019-105586](https://doi.org/10.1136/medethics-2019-105586)] [Medline: [31748206](https://pubmed.ncbi.nlm.nih.gov/31748206/)]
56. Hitzenko K, Cowan HR, Goldrick M, Mittal VA. Racial and ethnic biases in computational approaches to psychopathology. *Schizophr Bull* 2022 Mar 01;48(2):285-288. [doi: [10.1093/schbul/sbab131](https://doi.org/10.1093/schbul/sbab131)] [Medline: [34729605](https://pubmed.ncbi.nlm.nih.gov/34729605/)]
57. Mota NB, Pimenta J, Tavares M, Palmeira L, Loch AA, Hedin-Pereira C, et al. A Brazilian bottom-up strategy to address mental health in a diverse population over a large territorial area - an inspiration for the use of digital mental health. *Psychiatry Res* 2022 May;311:114477. [doi: [10.1016/j.psychres.2022.114477](https://doi.org/10.1016/j.psychres.2022.114477)] [Medline: [35245744](https://pubmed.ncbi.nlm.nih.gov/35245744/)]
58. Mota NB, Sigman M, Cecchi G, Copelli M, Ribeiro S. The maturation of speech structure in psychosis is resistant to formal education. *NPJ Schizophr* 2018 Dec 07;4(1):25. [doi: [10.1038/s41537-018-0067-3](https://doi.org/10.1038/s41537-018-0067-3)] [Medline: [30531913](https://pubmed.ncbi.nlm.nih.gov/30531913/)]
59. Walker M. Otto Hahn: responsibility and repression. *Phys Perspect* 2006 May;8(2):116-163. [doi: [10.1007/s00016-006-0277-3](https://doi.org/10.1007/s00016-006-0277-3)]
60. Zierler D. *The Invention of Ecocide: Agent Orange, Vietnam, and the Scientists Who Changed the Way We Think About the Environment*. Athens, GA: University of Georgia Press; 2011.
61. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. *JMIR Ment Health* 2022 Apr 14;9(4):e33450 [FREE Full text] [doi: [10.2196/33450](https://doi.org/10.2196/33450)] [Medline: [35436240](https://pubmed.ncbi.nlm.nih.gov/35436240/)]
62. Nicholas J, Onie S, Larsen ME. Ethics and privacy in social media research for mental health. *Curr Psychiatry Rep* 2020 Nov 23;22(12):84. [doi: [10.1007/s11920-020-01205-9](https://doi.org/10.1007/s11920-020-01205-9)] [Medline: [33225404](https://pubmed.ncbi.nlm.nih.gov/33225404/)]

## Abbreviations

**AI:** artificial intelligence



**ARMS:** at-risk mental state

**AUC:** area under the curve

**ICU:** intensive care unit

**ML:** machine learning

*Edited by J Torous; submitted 13.07.22; peer-reviewed by A Joseph, K Uludag; comments to author 19.08.22; revised version received 09.09.22; accepted 04.10.22; published 01.11.22*

*Please cite as:*

*Loch AA, Lopes-Rocha AC, Ara A, Gondim JM, Cecchi GA, Corcoran CM, Mota NB, Argolo FC*

*Ethical Implications of the Use of Language Analysis Technologies for the Diagnosis and Prediction of Psychiatric Disorders*

*JMIR Ment Health 2022;9(11):e41014*

*URL: <https://mental.jmir.org/2022/11/e41014>*

*doi: [10.2196/41014](https://doi.org/10.2196/41014)*

*PMID:*

©Alexandre Andrade Loch, Ana Caroline Lopes-Rocha, Anderson Ara, João Medrado Gondim, Guillermo A Cecchi, Cheryl Mary Corcoran, Natália Bezerra Mota, Felipe C Argolo. Originally published in JMIR Mental Health (<https://mental.jmir.org>), 01.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on <https://mental.jmir.org/>, as well as this copyright and license information must be included.