1-1-2020

"Sorry I Didn't Hear You." The Ethics of Voice Computing and AI in High Risk Mental Health Populations

Fazal Khan
Assistant Professor of Law *University of Georgia School of Law*, fkahn@uga.edu

Christopher Villongco
Assistant Professor of Law *Morehouse School of Medicine*

AJOB NEUROSCIENCE
2020, VOL. 11, NO. 2, 105–112
https://doi.org/10.1080/21507740.2020.1740355

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

"Sorry I Didn't Hear You." The Ethics of Voice Computing and AI in High Risk Mental Health Populations

Christopher Villongco[a] and Fazal Khan[b]

[a]Morehouse School of Medicine; [b]University of Georgia School of Law

ABSTRACT

This article examines the ethical and policy implications of using voice computing and artificial intelligence to screen for mental health conditions in low income and minority populations. Mental health is unequally distributed among these groups, which is further exacerbated by increased barriers to psychiatric care. Advancements in voice computing and artificial intelligence promise increased screening and more sensitive diagnostic assessments. Machine learning algorithms have the capacity to identify vocal features that can screen those with depression. However, in order to screen for mental health pathology, computer algorithms must first be able to account for the fundamental differences in vocal characteristics between low income minorities and those who are not. While researchers have envisioned this technology as a beneficent tool, this technology could be repurposed to scale up discrimination or exploitation. Studies on the use of big data and predictive analytics demonstrate that low income minority populations already face significant discrimination. This article urges researchers developing AI tools for vulnerable populations to consider the full ethical, legal, and social impact of their work. Without a national, coherent framework of legal regulations and ethical guidelines to protect vulnerable populations, it will be difficult to limit AI applications to solely beneficial uses. Without such protections, vulnerable populations will rightfully be wary of participating in such studies which also will negatively impact the robustness of such tools. Thus, for research involving AI tools like voice computing, it is in the research community's interest to demand more guidance and regulatory oversight from the federal government.

KEYWORDS

Computers; mental health; psychiatry; representations

INTRODUCTION: ASSESSING MENTAL HEALTH NEEDS THROUGH VOICE COMPUTING

Voice computing is defined as a discipline that aims to develop hardware or software to process voice inputs.[1] The utilization of voice computing has made a tremendous impact on our society today. In fact, with the development and release of Siri on Apple iPhones in addition to the Amazon Echo and Google Voice Assistant, we may have moved into the first voice-first period: A dramatic shift in how users interact with computers from a voice first perspective as opposed to interacting with the traditional screen, mouse, and keyboard (Schwoebel 2018). Voice computing not only promises advancements in the technology field, but in the field of Mental Health as well. Many psychiatric illnesses have a voice component to their diagnostic criteria. Whether it is disorganized speech as a key feature to make the diagnosis of schizophrenia, or symptoms of being more talkative than usual or pressure to keep talking being characteristic of manic episodes found in a disease such as Bipolar I and II (American Psychiatric Association 2013), voice computing promises to be a groundbreaking diagnostic tool. A study done in 2015 utilized automated speech analysis to look at transcribed interviews of youth deemed high risk to develop a psychotic episode. The utilization of this automated speech analysis was able to predict with 100% accuracy which individual with prodromal symptoms would ultimately transition into full blown psychosis (Bedi et al. 2015). Another study that compared acoustic features was able to discriminate between depressed and controlled patients with a sensitivity of 77.8% and a specificity of 86.1% on key vocal frequencies alone (Taguchi et al. 2018). When combining both linguistic analysis of transcriptions and acoustic features, Pestian et al. successfully distinguished between suicidal patients vs. control with 92% accuracy (Pestian et al. 2017) from a cohort of

CONTACT Christopher Villongco cvillongco@msm.edu Morehouse School of Medicine, Atlanta, GA 30310-1458, USA.

patients entering the emergency department. While voice computing boasts a very promising future as a diagnostic tool, the population which it will serve and how it will be used in clinical decision making has yet to be established.

POTENTIAL BENEFITS AND HARMS OF USING VOICE COMPUTING TO IDENTIFY VULNERABLE PATIENTS

One population which critically needs mental health services, in general, are minority patients and those with low socioeconomic status (low-SES). A meta-analysis found that just being categorized as low-SES alone already significantly increased one's odds of being depressed (odds ratio = 1.81) (Lorant et al. 2003). A 7-year prospective cohort study found that changes in financial strain, deprivation, and poverty led to significant increases in both depressive symptoms as well as increased clinically depressive episodes (Lorant et al. 2007). When looking at minority populations, it has been well established that compared to their white counterparts there is a racial disparity gap in lower access and utilization of mental health services. This disparity gap can be measured by comparing psychotropic drug use and expenditures, "which serve as a proxy for measuring adequate access to mental health care and services," among different racial and ethnic groups (Pierre et al. 2014). A study examining racial-ethnic trends in access to mental health care from 2004–2012 found that in the span of 8 years, the disparity gap between black-white psychotropic medication use increased. During this time, the disparity gap between hispanic-white in both mental health care utilization and psychotropic medication increased as well (Cook et al. 2017). When combining these 2 high risk populations, an even greater need for mental health services can be elicited. For example, one study looking at a heavily under-resourced community (only 22.5% currently employed and 64.5% having a family income of less than $10,000) working with a largely Latino and African American population in Los Angeles has a 33% greater chance of screening at least moderately depressed (Miranda et al. 2013). For affluent individuals who regularly have excellent access to their primary care physician (PCP) as well as an in-network psychiatrist, a diagnostic technology such as voice computing may not be as useful. This is because, while this technology can decrease the time it takes to be screened for mental health issues, they already have well-established care with their PCP and in turn, will have much less barriers to be diagnosed with and treated for psychiatric illness. Thus the mental health population who would most directly benefit from this technology are not the affluent but minority, low SES patients. Minority and low SES individuals do not have access or funds to be regularly screened and diagnosed with psychiatric illness so a tool that decreases potential barriers to this process would be of great benefit. However, in the wrong hands, this technology could be used to further discrimination and antagonism of an already disadvantaged population. To understand how this could occur, one must take a step back to understand the basic principles of how voice computing works.

HOW THE TECHNOLOGY WORKS

Voice computing as a way to address mental health at its core takes a statistical learning approach to the problem in the sense that one is using an input (different features of voice) to attempt to further explain a specific output (in this case a psychiatric illness or symptom). There are two major ways this statistical learning is accomplished: supervised learning and unsupervised learning (Gareth n.d). Briefly, in supervised learning—one feeds specific inputs (voice features) through a mathematical model that will predict or infer a specific output (psychiatric diagnosis). With unsupervised learning, one once again gives an input (voice features) but instead of trying to predict an outcome in a rigid way, the machine will associate these different features in order to explain the relationship that various inputs and outputs have with each other. Despite their differences, both statistical/machine learning methods have a couple key similarities. The first is that there must be a period for which a machine can learn before it can start making predictions or inferences (Schwoebel 2018). The second similarity, which is the more important concept for this paper, is the fact that there must be an actual training data set for which the machine must learn from—and that this training dataset is vital for the actual success of the machine in accomplishing its ultimate task.

BIASED DATA

While voice computing is still a developing field, other fields which utilized artificial intelligence have highlighted the dangerous consequences of not having a diverse dataset. Bolukbasi et al. looked at a predictive computer algorithm trained on a publicly available

dataset of Google News articles (Bolukbasi et al. 2016). Based on this training data set, the study found the program learned to incorporate male-female gender bias. In this study machine learning learned to associate words by learning which words co occur in the same text body. The machine learning model was found to start associating the word "she" with gender stereotyped occupations (housekeeper, homemaker, nanny). Another example of how unrepresentative data sets can lead to algorithmic discrimination was a study done by Caliskan et al. further looking at word embedding. Word embedding is a machine learning process which learns and predicts the associations between words that people use. Caliskan et al. found that their model which was trained of the World Wide Web, acquired prejudicial bias that reflected historical injustices experienced by African American populations (Caliskan et al. 2017). By measuring the vector difference between words, a machine learning algorithm learned how to attribute and predict which words should be associated with each other. Caliskan found the post training, AI were more likely to associate traditionally European American names (i.e., Adam, Nancy, etc.) with pleasant adjective (i.e., "health, love, peace") when compared to names considered traditionally African American (Leroy, Aiesha, Lakisha, etc.). Artificial Intelligence utilization in facial recognition serves as an important comparison for voice computing. A study looking at artificial intelligence working with mug shot face images from county sheriff's office found that the facial recognition algorithm had difficulty accurately recognizing and matching female, black and younger cohorts. The study also found that face recognition performance improved when training of the machine algorithm occurred exclusively on the specific cohort being addressed(Klare et al. 2012). These studies highlight the importance of diversity and careful analysis of the training data sets.

Machine learning and artificial intelligence are not implicitly biased, but if their data training sets contain culturally biased labeling, the model's outputs will reveal this. Consequently, once these specific data sets are collected, careful consideration must be taken at how these data sets are utilized. While there is an obvious need for diverse data sets, one must keep in mind that these are already very disadvantaged communities and populations. Historically and systemically, these communities have been biased against and if this technology is not utilized correctly, it could ultimately be a tool that widens the gap as opposed to its original purpose and design.

WORK IN PROGRESS: ETHICAL AND REGULATORY GUIDELINES FOR AI RESEARCH ON HUMAN SUBJECTS

The ethical challenges described above are clearly not unique to voice computing applications, as other types of AI research that leverage biometric information, such as facial recognition or genetic phenotyping, raise similar concerns (Wee and Mozur 2019). However, the case study above crystallizes the dilemma scientists face when particular applications of their research can be used in ways that can either benefit or significantly harm the individuals and populations being studied. Exacerbating this challenge is the underdevelopment of ethical and regulatory guidelines for this type of research, which is only growing in number and scale. This section will highlight these gaps and critically assess regulatory attempts by U.S. state and federal governments, as well as self-regulatory initiatives by Facebook and Google. Lastly, this section will argue that it is in the AI research community's interest for the federal government to promulgate more robust legal protections and ethical research guidelines than what the current Trump Administration has been proposing.

Historically, the pattern of development for human subject research guidelines has been reactionary as opposed to proactive. For instance, the Nuremburg Code was a response to the shocking revelations about the Nazi medical trials and the federal Belmont Report and Common Rule were a response to the Tuskegee Syphilis Trials. There is an understandable logic to this pattern–regulators do not want to get ahead of new areas of scientific inquiry for fear of stifling novel and beneficial innovations. As the influential law and economics scholar and jurist Richard Posner opined in one case, "Law lags science; it does not lead it" (*Rosen v. Ciba-Geigy Corp.*, 1996). However, as stated in a recent report by the International Council of Scientists, "scientists as individuals and the international scientific community have a shared responsibility, together with other members of society, to do their utmost to assure that scientific discoveries are used solely to promote the common good" (Paris: International Council for Science, 2014). This statement reflects a tacit social contract between the public and scientific research community that is mutually beneficial. However, when the scientific community ignores this proscription or exploits regulatory loopholes, a social backlash often follows. Indeed, as the public has become more concerned over the implications of big data and AI research, there is a growing recognition from within

the research community that they need clearer guidelines when it comes to this type of research.

The federal "Common Rule" is a collection of federal regulations that regulate human subjects research. Universities that receive federal funding for human research have to comply with these rules, which includes setting up institutional review boards (IRBs) to ensure compliance with the Common Rule. In addition, private companies that conduct human clinical trials for the purpose of developing pharmaceuticals, medical devices, or biologics, are also subject to oversight by the Food and Drug Administration (FDA). Yet, if a private company conducts research on human populations using digital, rather than biological samples, and they are not developing a product subject to FDA regulations, they are exempt from following the Common Rule and federal regulations. This presents a regulatory challenge, as more and more human research leverages AI to analyze social media and other online content (e.g., personal pictures and videos). Even among scholars in the research ethics community, there is a "consensus in favor of the status quo … [as] extending the Common Rule to all human subjects research would be cumbersome and impractical to enforce" (Relias Media 2014). However, this stance is based on the view that "most of the studies in question would be of low risk and more akin to surveys and quality improvement than clinical trials" (Relias Media 2014). This characterization of most of these studies being "low risk" makes sense if one is applying the clinical trials paradigm of minimizing physical or mental harm to subjects, but what if the risk of harm from a study comes in the form of damage to one's political or economic rights? The problem is that the traditional Common Rule framework is ill-suited to consider these other types of harms.

There has been development in the research ethics literature for "dual-use research of concern (DURC)" and a growing body of normative practices to mitigate potential harms from such research. The imperative for developing DURC regulations and research guidelines became salient following the anthrax attacks in Washington, D.C. that paralyzed the nation's capital only weeks after the terrorist attacks on 9/11. However, work in this area has almost exclusively occurred within the context of research on pathogens, including viruses and bacteria, that implicate biosafety concerns. As with the Common Rule, there has been no push to extend DURC-like oversight and precautions to AI studies, such as publication or research restrictions based on risk concerns.

INDUSTRY SELF-REGULATION

Facebook's now infamous "emotional contagion" research trial illustrates the regulatory gaps for big-data research conducted by private corporations (Kramer et al. 2014). In this 2014 study, Facebook ran an experiment on over 680,000 of the social media site's users, manipulating a user's "News Feed" to show either predominantly happy or sad content. Then Facebook measured if the emotional content of a user's News Feed had an impact on that subjects' mood. The study reported that there was an emotional contagion effect in social media, as users who had a "happier" news feed tended to post more emotionally positive content, while those with a "sadder" news feed tended to post more emotionally negative content. The subsequent publication of this study provoked outrage as it revealed that Facebook was experimenting with its users emotional state without informing them. Additionally, it is plausible that this intervention could cause more than "minimal harm" to those who were susceptible to or already suffering from depression. Interestingly, while a Cornell University researcher analyzed data sets from this study, the Cornell IRB determined that this did not constitute "human research" as he only had access to de-identified data and was not involved in the initial study intervention by Facebook.

Facebook responded that it did nothing wrong and in fact did receive consent from users to do these types of studies as part of its contractual terms of service that its users accepted by clicking a box online. Within these terms of service, Facebook highlighted its "Data Use Policy" which described how it might use its customers information, including "for internal operations, including troubleshooting, data analysis, testing, research and service improvement." Of course, the unsatisfying nature of Facebook's explanation is that the consent they describe is rarely informed or meaningful, as users invariably scroll through thousands of words without reading them in order to reach the clickable box that will activate their account (Kramer et al. 2014). Thus, while Facebook could plausibly argue that this study violated no legal regulations or contractual terms, it seemed to recognize that this argument failed to address a serious ethical concern by violating an implicit social contract with its over 2.3 billion users.

As a response to the public backlash highlighted above, many large technology companies like Facebook and Google have announced self-regulatory efforts, including setting up internal review boards to assess the ethical implications of their research

projects. Notably, the purview of these review boards go beyond the protection of human research subjects, to "touch on the broader question of whether the insights gained from the research might have harmful downstream consequences on a wider population" (Hartzog 2011; MacCarthy 2019). However, both of these companies illustrate how difficult ethical self-regulation can be when significant economic conflicts exist. For example, subsequent to the creation of Facebook's internal review board, Cambridge-Analytica was able to use the "scraped" Facebook profiles of 87 million users without their consent in order to create "pyschographic" profiles of U.S. voters (Rosenberg et al. 2018). In turn, there is overwhelming evidence that Cambridge-Analytica and the Russian government used Facebook to target voters in the 2016 elections with highly divisive political messages, including content intended to stoke anti-immigrant and racial hostilities (Isaac and Shane 2017).

In April 2018, Google faced strong internal dissent from thousands of its employees, who signed a letter protesting the company's development of AI technology for improving the efficacy of military drone strikes (Wakabayashi and Shane 2018). Consequently, Google's leadership responded a few months later to this internal challenge by dropping its military contract and publicly announcing a set of seven core AI principles that would guide its development and use of AI technology: (1) Be socially beneficial; (2) Avoid creating or reinforcing unfair bias; (3) Be built and tested for safety; (4) Be accountable to people; (5) Incorporate privacy design principles; (6) Uphold high standards of scientific excellence; and (7) Be made available for uses that accord with these principles. Google further explained that it would not pursue development of AI applications under the following circumstances:

> Where we believe that the benefits substantially outweigh the risks and… that gather or use information for surveillance violating internationally accepted norms… [or] whose purpose contravenes widely accepted principles of international law and human rights. (Sterling 2018)

While many initially praised what looked like a principled response by Google to elevate the ethical concerns raised by its employees over pure economic profit, this response has been tempered by subsequent reports indicating that Google is finding other indirect ways to support development of military AI technology through its own venture capital subsidiary called Gradient Ventures. Through Gradient Ventures, Google is "providing financial, technological, and engineering support" to a host of startups that are developing AI technologies for military and law enforcement purposes (Fang 2019). Thus, as dissenting Google employees have asserted, the company is circumventing its own self-imposed ethical commitments and standards. The lesson perhaps is that industry's adherence to voluntary guidelines in the development of ethical AI tools is a weak and inconstant form of regulation.

REGULATION OF AI TECHNOLOGY AT THE STATE AND FEDERAL LEVEL

AI ethics scholar Rodrigo Ochigame provides a useful framework for assessing three regulatory possibilities for a given technology: "(1) no legal regulation at all, leaving 'ethical principles' and 'responsible practices' as merely voluntary; (2) moderate legal regulation encouraging or requiring technical adjustments that do not conflict significantly with profits; or (3) restrictive legal regulation curbing or banning deployment of the technology" (Ochigame 2019). As expected, the technology industry has largely supported the first two regulatory options while vociferously opposing more stringent regulations. While there has been little direct regulatory action by the federal government on potentially harmful uses of AI-dependent technologies, in recent years states like Illinois, Washington, Texas, and California have either passed or proposed biometric privacy laws. A key feature if these laws is to prohibit the use of biometric information for commercial purposes without an individual's consent. Illinois' law, in particular, has been a concern for industry as it creates a private right of action (i.e., allowing individuals to sue for violations of the law) for technical violations of the law and does not require a plaintiff to show actual damages. Consequently, in 2020 Facebook paid $550 million to settle a class-action lawsuit over alleged misuse of facial-recognition technology for affected users within Illinois (Singer and Isaac 2020). As other states pass or are considering passing similar biometric privacy laws, this can obviously send a strong regulatory deterrence signal to the technology industry through the tort system.

In January 2020, however, the Trump Administration signaled through a draft memorandum issued by the Office of Management and Budget (OMB) on "Guidance for Regulation of Artificial Intelligence Applications ("OMB AI Memorandum") that it is aligned with industry's preference for a relaxed regulatory environment. The OMB AI Memorandum, created in conjunction with the Directors of the Office and Science and Technology

Policy, Domestic Policy Council, and National Economic Council, illustrates the AI regulatory preferences of the Trump Administration:

> Federal agencies must avoid regulatory or non-regulatory actions that needlessly hamper AI innovation and growth. Where permitted by law, when deciding whether and how to regulate in an area that may affect AI applications, agencies should assess the effect of the potential regulation on AI innovation and growth. Agencies must avoid a precautionary approach that holds AI systems to such an impossibly high standard that society cannot enjoy their benefits. Where AI entails risk, agencies should consider the potential benefits and costs of employing AI, when compared to the systems AI has been designed to complement or replace. Furthermore, in the context of AI, as in other settings, agencies must consider the effect of Federal regulation on existing or potential actions by State and local governments. In some circumstances, agencies may use their authority to address inconsistent, burdensome, and duplicative State laws the prevent the emergence of a national market. Where a uniform national standard for a specific aspect related to AI is not essential, agencies should consider forgoing regulatory action. (Executive Order No. 13859 2019)

From a regulation and policy perspective, two major themes stand out in the OMB AI Memorandum. First, Federal regulations on AI technologies are disfavored and agencies have to meet a high threshold to justify regulatory and non-regulatory (e.g., guidelines and recommendations) actions in this sphere. Second, Federal agencies are encouraged to use their authority to preempt State laws that are seen as too burdensome for the deployment of AI technology.

In our analysis, the policy framework of the OMB AI Memorandum does not take the nation in the right direction for regulating AI technologies and promoting trust that they will be beneficial for the populations that they are deployed on. The laissez-faire approach promoted by this document may actually frustrate the goals of promoting innovation and growth in AI technologies, as lack of trust and unchecked risks from such research can lead to a backlash by the public and potential research subjects. Beyond this concern over AI innovation, the democratic, civil, and human rights of vulnerable populations should not be sacrificed in the name of technological progress.

MENTAL HEALTH SMARTPHONE APPLICATIONS: DRAWING THE LINE AT SAFETY

The most likely utility of voice computing in the field of mental health would be used through a smartphone app. Thus parallels in smartphone technology can guide how voice computing should be regulated when dealing with mental health issues. Although the FDA has already been discussed, the U.S Federal Trade Commission (FTC) is another governing body which regulates software that is used in medical diagnosis and treatment. In 2017, the FTC filed a complaint against a Breathometer, Inc—which created an app promising to result an individual's blood alcohol concentration (BAC) using data collected through a device and analyzed in a smartphone (Federal Trade Commission 2017). The complaint settled and in a statement by the commissioner stated, "The complaint alleges that Breathometer made false and unsubstantiated advertising claims that their breathalyzer devices had "undergone rigorous government lab grade testing" for ability to accurately detect consumers' blood alcohol content (BAC) and were "law-enforcement grade product[s]" for the purpose of complying with impaired driving laws.1 I support this complaint and settlement. Companies must substantiate their advertising claims, and I have reason to believe that Breathometer failed to do so" (Federal Trade Commission, 2017). While no voice computing has been utilized in smartphone applications for mental health—there are similarities we can draw. Both Breathometer and voice computing promise similar strategies, using smartphone metrics to help in mental health diagnosis. The case which eventually settled showed that the government will regulate if there is a potential for harm that is caused by falsely marketing the true utility of a smartphone app. If future industries start utilizing recorded metrics (such as voice computing) for uses other than what is advertised, the FTC could follow a similar process in order to ultimately protect low income minorities if they believe that the misutilization of the digital metrics recorded are being recorded and utilized for uses other than what was marketed, ultimately leading to detrimental outcomes for the original user.

UTILIZING VOICE COMPUTING TO BENEFIT THE COMMUNITY

Moving forward, two main principles that must be considered are community engagement and equal allocation. The importance of multidisciplinary community engagement in AI technology is highlighted in the Jackson Heart Study. This study investigated the major genetic and environmental risk factors which contributed to the disproportionately poor cardiovascular outcomes of African Americans in Jackson,

Mississippi (Sempos et al. 1999). A component of this study utilized machine learning to detect previously unseen associations between hypertension in this high risk population (Seffens et al. 2016) as well as developing mobile health platform for data collection and educational resources for high risk African Americans (Taylor et al. 2018). Due to the multifactorial nature of hypertension, researchers could not rely solely on previously established metrics of cardiovascular health but had to work closely with multidisciplinary teams to engage the community. Their major objectives included taking time to learn historical and cultural aspects of the community as well as establishing present and future value to the community. In a similar vein, equal allocation is a process in which one ensures resources are proportionally allocated to the individuals which are algorithmically determined to need the resources the most (Rajkomar et al. 2018). This is completed by taking into account baseline differences between groups and calibrating for this bias before one runs the metrics through machine learning algorithms (Pleiss et al. 2017).

Much like hypertension, psychiatric illness is a multifactorial disease that requires a comprehensive understanding of not only measurable biometric data, but the complex social and historical context which also contribute heavily to outcomes. Much like the Jackson Heart Study, those hoping to utilize voice computing in mental health must collaborate heavily with the community data is collected from to ensure that the data is addressing the needs of community as well as improving mental health outcomes in that community. This will be accomplished by utilizing a multidisciplinary team who heavily understand low SES minority populations. Further building on this the principle, equal allocation will ensure that once deficits and goals of the community are found, these differences are calibrated for when utilizing AI.

The utilization of voice computing is in the research phase and not fully utilized in the market for healthcare treatment yet, however, abiding by these principles will ensure that this AI technology completes its originally proposed goal: improved mental health wellness in a community which historically experienced increased barriers to care.

CONCLUSION

As described above, voice computing and other AI applications have great potential to benefit individual and population health, by making needed care more accessible and affordable. However, it is also clear that the AI research community, including in universities and private corporations, needs more ethical guidance and regulatory guardrails when such technology could be used in manners that are both beneficial and harmful to particular individuals and groups in society. Further, given that AI research is qualitatively different than the human subject research contemplated by the Common Rule, it makes sense to create a novel ethical and regulatory framework for AI technologies used on humans. To push this effort forward, the research community, private industry, and the federal government need to come together, to create an AI-specific Common Rule to help ensure that the social contract between the science community and the public remains intact.

DISCLOSURE STATEMENT

Dr. Villongco works as a research intern at a Neurolex Laboratories, a voice computing company. He has not and does not receive any financial compensation or benefits.

REFERENCES

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: Author.

Bedi, G., F. Carrillo, G. A. Cecchi, et al. 2015. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia* 1(1): 15030. doi: 10.1038/npjschz.2015.30.

Bolukbasi, T., K. W. Chang, J. Zou, V. Saligrama, and A. Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems* 29.

Caliskan, A., J. J. Bryson, and A. Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334): 183–186. doi: 10.1126/science.aal4230.

Cook, B. L., N. H. Trinh, Z. Li, S. S. Hou, and A. M. Progovac. 2017. Trends in racial-ethnic disparities in access to mental health care, 2004–2012. *Psychiatric Services* 68(1): 9–16. doi: 10.1176/appi.ps.201500453.

Executive Order No. 13859. 2019. 84 C.F.R 3967 (February 14).

Fang, L. 2019. Google continues investments in military and police AI technology through venture capital arm. *The Intercept*, July 23. Retrieved from https://theintercept.com/2019/07/23/google-ai-gradient-ventures/

Federal Trade Commission. 2017. Concurring Statement of Commissioner Maureen K. Ohlhausen In the Matter of Breathometer, Inc. Press release, Janurary 23. Retrieved from https://www.ftc.gov/system/files/documents/public_statements/1054953/170123breathometerohlhausenstatement.pdf

Gareth, J. n.d. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer Verlag.

Hartzog, W. 2011. Website design as contract. *60 American University Law Review 1635*. Retrieved from https://ssrn.com/abstract=1808108

International Science Council. 2014. Freedom, responsibility and universality of science. August 29, 2019. Retrieved from https://council.science/publications/freedom-responsibility-and-universality-of-science-2014/

Isaac, M., and S. Shane. 2017. Facebook's Russia-linked ads came in many disguises. *New York Times*, October 2. Retrieved from https://www.nytimes.com/2017/10/02/technology/facebook-russia-ads-.html

Klare, B. F., M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain. 2012. Face recognition performance: Role of demographic information. *IEEE Transactions on Information Forensics and Security* 7(6): 1789–1801. doi: 10.1109/TIFS.2012.2214212.

Kramer, A. D., J. E. Guillory, and J. T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of Sciences* 111(24): 8788–8790. doi: 10.1073/pnas.1320040111.

Lorant, V., C. Croux, S. Weich, D. Deliège, J. Mackenbach, and M. Ansseau. 2007. Depression and socio-economic risk factors: 7-year longitudinal population study. *British Journal of Psychiatry* 190(4): 293–298. doi: 10.1192/bjp.bp.105.020040.

Lorant, V., D. Deliège, W. Eaton, A. Robert, P. Philippot, and M. Ansseau. 2003. Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology* 157(2): 98–112. doi: 10.1093/aje/kwf182.

MacCarthy, M. 2019. How to address new privacy issues raised by artificial intelligence and machine learning. *Brookings*, June 13. Retrieved from https://www.brookings.edu/blog/techtank/2019/04/01/how-to-address-new-privacy-issues-raised-by-artificial-intelligence-and-machine-learning/

Miranda, J., M. K. Ong, L. Jones, et al. 2013. Community-partnered evaluation of depression services for clients of community-based agencies in under-resourced communities in Los Angeles. *Journal of General Internal Medicine* 28(10): 1279–1287. doi: 10.1007/s11606-013-2480-7.

Ochigame, R. 2019. How big tech manipulates academia to avoid regulation. *The Intercept*, December 20. Retrieved from https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

Pestian, J. P., M. Sorter, B. Connolly, et al. 2017. A machine learning approach to identifying the thought markers of suicidal subjects: A prospective multicenter trial. *Suicide and Life-Threatening Behavior* 47(1): 112–121. doi: 10.1111/sltb.12312.

Pierre, G., R. J. Thorpe, G. Y. Dinwiddie, and D. J. Gaskin. 2014. Are there racial disparities in psychotropic drug use and expenditures in a nationally representative sample of men in the United States? Evidence from the Medical Expenditure Panel Survey. *American Journal of Men's Health* 8(1): 82–90. doi: 10.1177/1557988313496564.

Pleiss, G., M. Raghavan, F. Wu, J. Kleinberg et al. 2017. On fairness and calibration. In *Proceedings from the Conference on Advances in Neural Information Processing Systems 2017, Long Beach, California, 4–9 December 2017*, eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, and S. Vishwanathan, 5680–5689. La Jolla, CA: Neural Information Processing Systems.

Rajkomar, A., M. Hardt, M. D. Howell, G. Corrado, and M. H. Chin. 2018. Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine* 169(12): 866–872. doi: 10.7326/M18-1990.

Relias Media. 2014 The Facebook Study and the common rule. Retrieved from https://www.reliasmedia.com/articles/118880-the-facebook-study-and-the-common-rule

Rosen v. Ciba-Geigy Corp., 78 F.3d 316. (7th Cir., 1996).

Rosenberg, M., N. Confessore, and C. Cadwalladr. 2018. How trump consultants exploited the Facebook data of millions. *New York Times*, March 17. Retrieved from https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Schwoebel, J. 2018. *An introduction to voice computing in python*. Vol. 1. Boston, MA: Neurolex Labratories.

Seffens, W., C. Evans, Taylor Minority Health-GRID Network, and Herman. 2016. Machine learning data imputation and classification in a multicohort hypertension clinical study. *Bioinformatics and Biology Insights* 9(Suppl 3): 43–54. doi: 10.4137/BBI.S29473.

Sempos, C. T., D. E. Bild, and T. A. Manolio. 1999. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. *The American Journal of the Medical Sciences* 317(3): 142–146. doi: 10.1016/S0002-9629(15)40495-1.

Singer, N., and M. Isaac. 2020. Facebook to pay $550 million to settle facial recognition suit. *New York Times*, January 29. Retrieved from https://www.nytimes.com/2020/01/29/technology/facebook-privacy-lawsuit-earnings.html

Sterling, B. 2018. Google's AI principles. *Wired Magazine*, June 9. Retrieved from https://www.wired.com/beyond-the-beyond/2018/06/googles-ai-principles/

Taguchi, T., H. Tachikawa, K. Nemoto, et al. 2018. Major depressive disorder discrimination using vocal acoustic features. *Journal of Affective Disorders* 225: 214–220. doi: 10.1016/j.jad.2017.08.038.

Taylor, H. A., F. Henderson, A. Abbasi, and G. Clifford. 2018. Cardiovascular disease in African Americans: Innovative community engagement for research recruitment and impact. *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 72(5): S43–S46. doi: 10.1053/j.ajkd.2018.06.027.

Wakabayashi, D., and S. Shane. 2018. Google will not renew pentagon contract that upset employees. *New York Times*, June 1. Retrieved from https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html

Wee, S.-L., and P. Mozur. 2019. China uses DNA to map faces, with help from the west. *New York Times*, December 3. Retrieved from https://www.nytimes.com/2019/12/03/business/china-dna-uighurs-xinjiang.html