JUST TECH

FIELD REVIEW

Big Bad Feelings: AI Depression Diagnosis and the Technopolitics of Disability

By: Jeff Nagy

DISCO

Nagy, Jeff. "Big Bad Feelings: AI Depression Diagnosis and the Technopolitics of Disability." Just Tech. Social Science Research Council. February 7, 2024. DOI: https://doi.org/10.35650/JT.3065.d.2024.

ABSTRACT

At a time when human clinicians are in short supply, AI tools promise access to mental health care via NLP chatbots, mindfulness apps, and adaptive CBT workbooks. But AI is increasingly also targeted upstream from treatment, at big-data-based diagnosis. For instance, Project Seabreeze, a joint effort between Apple and UCLA, aims to use passively-collected data from iPhones and Apple Watches to diagnose latent depression. This field review addresses systems like Project Seabreeze as an opportunity to think at the intersection of depression, disability, and digital media, with an eye to the logics of translation and scale. Ten years ago, Ann Cvetkovich described depression as a "public feeling," situating what she calls "feeling bad" as a defining state of life under conditions of neoliberalism, the cumulative result of entangled systems of disenfranchisement and oppression. Tools like Seabreeze transmute depression from a "public feeling" to a one diagnosed at the interface between platform-based big data and the fluid "scalable subjects" (Stark 2018) produced by algorithmic psychometrics. What happens to the political economy of diagnosis in this scalar shift, and in the concomitant translation from DSM-5 diagnostic criteria as understood and used by clinicians to a set of machine learning features? How does the algorithmization of diagnosis crystallize a condition whose only constant is its ever-changing definition (Ehrenberg 2010)? This field review takes up these questions, while considering what this emergent technopolitics of depression might mean for the politics of disability writ large.

I received my diagnosis the old-fashioned way: from a psychiatrist, in the course of a clinical evaluation. After an introductory session during which I held forth on what I then saw with grim certainty as the deep and unbearably painful meaninglessness of my life, the doctor asked me a series of questions: Would I say I was sleeping more than normal, or less? Was I having trouble concentrating on work or enjoying what I typically might? Was I spending a significant amount of time ruminating on past sins? Did

I not eat enough or too much or move or speak less or more slowly than normal? Was I thinking about death, or had I fantasized about killing myself? How long would I say all of these things had been going on? Two weeks? Two months? Two years? Longer?

I answered these questions and others, and at the end of the session I asked him for the verdict. To which he responded, more avuncularly than seemed totally appropriate, “I’d say it looks like a classic case of major depression.” From then on, my monthly invoices and the claims submitted to my insurance, like those of millions of others both in the United States and abroad, would bear the short string of characters that corresponds to a diagnosis of Major Depressive Disorder as defined in the American Psychiatric Association’s *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition, Text Revision (the DSM-5-TR).

This process has played out in more or less the same fashion between millions of patients and hundreds of thousands of clinicians over the past handful of decades. But if you’re imagining me lying on a couch, box of tissues conveniently nearby, and watched over by a bust of Freud or Kraepelin, there was one less-than-traditional aspect to the scenario. Since all of this took place at the height of the pandemic, I was diagnosed in the same digital space where I saw my friends, taught my classes, and presented research at conferences: Zoom. Years later, although I see my psychiatrist twice a week for what my invoices describe as “medication management” and “psychotherapy,” I have never met him in person, never thumbed through the yellowing copies of the *New Yorker* I imagine in his New Haven waiting room, never avoided eye contact with the appointment immediately before or after my own. I have never even seen his legs.



Spooky Diagnosis at a Distance

In Philadelphia, a Black woman in her early thirties is diagnosed with depression on the basis of the words she uses in her Facebook status updates (Eichstaedt et al. 2018). In Burlington, a five-year-old boy is diagnosed according to shifts in the pitch of his voice as he improvises a story for a stern adult (McGinnis et al. 2019). In Los Angeles, a middle-aged man is flagged as at risk for the condition from the way he uses his iPhone and Apple Watch, triangulating data about sleep, movement, heart rate, and online activity (Kisliuk 2020; Winkler 2021). Elsewhere, still others are diagnosed by a variety of passively-detected digital signals, everything from the aesthetic features of their Instagram photos (Reece and Danforth 2017), to the semantic content of their tweets (Tsugawa et al. 2015), to the emotional "fingerprint" of their posts and comments on subreddits (Guo et al. 2021).

In these cases, and many dozens of others like them, AI enables a spooky kind of diagnosis. Like quantum entanglement, it operates at a distance, invisibly in the background, without human intervention, and on a basis we sometimes barely or don't understand. Many of these systems are pilot projects and proofs-of-concept carried out by teams of scholars in psychiatry departments, schools of data science, or university hospitals. Their study populations are—usually—informed and consenting, and their methodologies and findings are published in journals and conference proceedings. But others, like the collaboration between UCLA and Apple, called Seabreeze, are undertaken in direct cooperation with corporate partners, who envision their own uses for digital phenotypes of psychiatric conditions. And the boundary between academic proof-of-concept and psychiatric industry is highly permeable, which has implications for a technopolitics of disability that stretches beyond these research clusters and study populations.

The past few years have seen an explosion of new digital systems targeted at mental health, a development accelerated by the Covid-19 pandemic. These go beyond simply replacing the clinician's consulting room with a Zoom grid. At a time when human clinicians are in short supply, AI tools promise access to mental health care via NLP (natural language processing) chatbots (Browne 2022), mindfulness apps (Jablonsky 2022), and adaptive CBT (cognitive behavior therapy) workbooks (e.g., Inkster, Sarda, and Subramanian 2018). But, in addition to these therapeutic innovations, AI is increasingly targeted upstream from treatment, at big data–based diagnosis, where it is meant to serve as an "early warning system" for everything from schizophrenia, to autism, to my own diagnosis, depression. Much of this research is framed as responding to an urgent need. Recognizing an accelerating and costly epidemic of undetected and chronic distress, AI aims to "cut the burden" (Quilantan 2018) by making diagnosis fast, cheap, and ubiquitous. Given the scale of the problem, it's often positioned as the only viable solution.

> *AI systems are now rewriting the relationship between depression, disability, digital media, and scale.*

There is, of course, a long history of "doctors who weren't there" (Greene 2022), delivering diagnoses and treatment through a range of technologies. Psychiatry is no different. Until recently, we may have thought of consultations and therapy as things that happened primarily in person, but technological mediation has been at its core since the days of Freud (Zeavin 2021). And, AI systems are now rewriting

the relationship between depression, disability, digital media, and scale. When AI relies on passively-monitored social media data, it follows in the footsteps of platform giants. Those companies' internal research teams have long documented the contributions their products make to depression (e.g., Wells, Horowitz, and Seetharaman 2021), while also touting their ability to pinpoint users' moods to investors and advertisers, leveraging similar AI systems to marketize user's psychic lives (Levin 2017).

A decade ago, Ann Cvetkovich (2012) described depression as a "public feeling," situating what she calls "feeling bad" as a defining state of life under conditions of disenfranchisement and oppression (Cvetkovich and Wilkerson 2016). AI diagnosis transmutes depression from a "public feeling" to a "social" one, detected at the interface between big data and the fluid "scalable subjects" (Stark 2018) produced by algorithmic psychometrics on social media platforms. What happens in this scalar shift, and in the concomitant translation from DSM-5 diagnostic criteria as understood and used by clinicians to a set of machine learning features? How do these tools, by bringing AI to bear on mental health, shift the political economy of diagnosis and the politics of disability?

AI Has Always Been Psychiatric

While these projects are often presented as the flower of an entirely novel interdisciplinarity, one made possible by AI's final maturity after decades of disappointment, AI depression detection is not the first time that psychiatry and computing have come together. In fact, their entanglement goes back to the very beginning of modern computing. Although the cybernetics movement that laid the groundwork for the development of artificial intelligence is now often remembered as a bleak postwar science of command and control, the models of the human and machine mind that cybernetics proliferated were shaped by continuous exchange with the psychiatry of the 1930s, 1940s, and 1950s (Pickering 2010; Nagy 2022). For instance, Warren McCulloch, cocreator with Warren Pitts of the Pitts-McCulloch neuron, an ancestor of contemporary neural networks, had trained as a psychiatrist, and spent the early 1930s working at what was then called the Rockland State Hospital for the Insane (Abraham 2016). He would later suggest that the more intelligent we made our machines, "the more surely they will have neuroses" (McCulloch 1949). In other words, as we made them smart enough to diagnose us, they too might become certifiable. The cyberneticians well knew, of course, that early models for neural networks were suggested by a psychiatrist's conception of the reverberating thought patterns of a human neurotic (Kubie 1930).

Early AI took the consulting room with it: The first chatbots were meant to imitate not William Shakespeare or Werner von Braun but a Rogerian analyst and a schizophrenic patient (Wilson 2010; Turkle 2005). In the 1950s, the same state psychiatric hospital where McCulloch had walked the wards two decades earlier was an early adopter of computing, with a center devoted to "psychoelectronics," using the mid-century version of big data to track, diagnose, and treat patients, and to test new drugs in large-scale trials (Nagy 2022). The biomedicalized way we currently think about depression comes directly out of this history. It was largely at Rockland, and through the efforts of its research director and computing division, that depression was recoded as the product of misaligned brain chemicals, treatable by psychopharmaceuticals—with patient histories, drug efficacy, and diagnostics all brought what they then styled "on-line."



Current building of the Rockland State Hospital for the Insane, now called the Rockland Psychiatric Center. Photo credit: Acroterion/Wikimedia Commons.

The Twin Epistemologies of Diagnostics and Big Data

If the history of AI shows that it emerged at least in part from the clinic, at their core, psychiatry and big data are natural bedfellows. They share an epistemological orientation that long precedes the recent emergence of digital diagnostic tools: Both abandon questions of underlying truth and causality in the search for reliable correlations. In big data, this is enough of a commonplace to warrant a catchphrase, "the end of theory" (Anderson 2008). Where a more traditional approach to knowledge production might posit a theory of causal relationships, make predictions based on that theory, and test those predictions experimentally, big data mining sieves enormous datasets for latent patterns. These patterns can be exploited without any curiosity about the causal relationships that produce them—no theory necessary. There's no need to have a theory as to why green banner ads for tax preparation services are more likely to be clicked than purple ones; you only need to know that they *reliably are* more clicked. Or why people will pay more for pad thai on Tuesdays in Peoria, only that they *reliably will*.

For psychiatry's part, the discipline has long had a troubled relationship with the ground truth of mental illness, and even to what "mental illness" might mean or be. As early as 1866, psychiatrists bemoaned the fact that, without a theory of the underlying nature of mental phenomena, "we are forced to fall back upon the symptomatology of the disease" (Greenberg 2013, 28). Across the twentieth century, from Freud down through the generations of his many warring disciples, the grand tradition of psychoanalysis sought the roots of psychiatric disease in various forms of psychic conflict. This theoretical tradition, once dominant even in hard-nosed and pragmatic American psychiatry, began to lose its grip over the course of the 1970s. Where clinicians and diagnostic manuals once genuflected in the direction of psychoanalytic concepts, by the end of the decade, they'd returned to a symptom-driven, theory-neutral framework. With the 1980 publication of the DSM-III, symptomatology became and still remains the

guiding principle for diagnostic categories. Speculation about underlying psychic causes was replaced by attempts to describe clusters of symptoms as closely and precisely as possible, with the goal of ensuring that different clinicians would arrive at the same diagnostic conclusion (Ehrenberg 2010, 154). But this emphasis on *reliability*—that similar clusters of symptoms would attract the same diagnosis across a population no matter the clinician—came at a cost. Reliability perhaps improved: No longer would one clinician see bipolar disorder where another would diagnose borderline personality disorder. But the search for *validity*—the way a diagnosis might be mapped on to a psychic ground truth—was largely abandoned.

> *The "end of theory" implicit in big data mining mirrors the atheoretical turn in modern diagnostic systems.*

In the portfolios of Silicon Valley venture capitalists, mental health is often positioned as an AI problem domain like routing self-driving cars, folding proteins for designer drugs, or tagging beheadings on YouTube. But, as I've just argued, the connection between AI and psychiatry runs deeper. The "end of theory" implicit in big data mining mirrors the atheoretical turn in modern diagnostic systems. From the DSM-III of the 1980s to today's DSM-5-TR, a diagnosis is not a handle on a structure of feeling; it is simply whatever clinicians reliably label or, in the context of a drug-driven model of depression, whatever *selective* serotonin reuptake inhibitors (SSRIs) seem to *selectively* treat (Davies 2015, 166). From this perspective, contemporary diagnosis is already a black box, one that nests quite naturally inside the larger black box of AI.

Feeling Bad: From the DSM to AI Diagnostics

Depression was the first target for big data–based psychiatric diagnosis (Chancellor et al. 2019). These efforts emerged from two strands of prior work in the early 2010s. First came the merger of data science and public health in "infodemiology," which claimed to predict outbreaks of epidemics like the annual flu from digital signals like Google searches (e.g., Ginsberg et al. 2009). Shortly thereafter, social psychology research began to track population-level variation in mood from aggregated posts on social media sites (e.g., Dodds et al. 2011; Kramer 2010). Like the flu, depression has repeatedly been positioned as an epidemic, a rhetoric routinely deployed to motivate AI interventions (e.g., Garg and Glick 2018). After all, if it was an epidemic like any other, and if scientists had already demonstrated that bad feelings could flow "contagiously" across social media (Kramer, Guillory, and Hancock 2014), we might leverage "infodemiological" tools to get a handle on this dramatic uptick of depression.

But it wasn't always this way. A relatively rare diagnosis in the 1940s and 1950s, depression only became the "common cold of psychiatry" in the 1970s, with a 1974 study by the World Health Organization (WHO) finding that a full fifth of the Western world had depressive symptoms (Harrington 2019, 202). What might explain this sudden rise to prominence and prevalence is debated. Did the growth of outpatient psychiatry catch depressions that didn't land their sufferers in psychiatric emergency rooms, or did the introduction of antidepressants give clinicians a new incentive to label patients in ways that

permitted prescribing (Sadowsky 2020, 97)? Were new checklists that collapsed subtypes and lowered the diagnostic bar to blame (Harrington 2019, 203)? Or was there something more metaphysical afoot, with depression's new centrality indexing a tectonic shift in psychiatry away from the delusions that characterized psychosis and toward questions of impaired affect (Ehrenberg 2010, 44)?

If, by the 1970s, psychiatrists knew they had an epidemic of *something*, that didn't mean they knew what exactly they had an overwhelming tide *of*. Across its history, the only certainty about depression is its confused and ever-changing definition. As Ehrenberg (Ehrenberg 2010, xxix) puts it, depression "brings together such a diversity of symptoms that the difficulty of defining and diagnosing it is a constant fact of psychiatry." Is it, as Freud thought, anger turned inward, or as his disciple Sandor Radó envisioned, a preemptive self-punishment for a too-strong reliance on the love of others? Or is it, as Aaron Beck, the father of cognitive behavioral therapy, argued, the result of faulty and self-defeating reasoning (Sadowsky 2020, 79)? Not even sadness is a constant focus: As antidepressants with a disinhibiting effect hit the market, psychiatrists came to see depression more as a pathological aversion to action more than as an insurmountable downer (Ehrenberg 2010, 141).

Even among the black boxes of psychiatric diagnosis codified in the DSM, depression might be a box blacker than others. Perhaps this epistemological muddle around depression's ground truth, which made it both a "master diagnosis" (Ehrenberg 2010, 1) and as common as the sniffles, also made it particularly amenable to AI prediction. Even if these epistemologies—of big data mining, of diagnostic categories, and of depression specifically—seem made for each other, funny things happen when symptom checklists are subsumed into machine learning features.



To see these shifts in action, let's consider the mainstream approach to AI depression detection. This approach uses machine learning to automatically sift large quantities of data for signals that might indicate depression. Unlike in deep learning approaches, discussed briefly below, for machine learning systems researchers build sets of features for their AI diagnosticians to attend to. Researchers take the diagnostic criteria for depression laid out in the DSM and find analogues that might be detected in the data at their disposal. If one symptom of depression is "sleep disturbance (insomnia or hypersomnia)," researchers might examine the timestamps of users' tweets to see if they're suddenly tweeting deep into the nocturnal void or, alternately, if they're mutely asleep much later into the day than normal (De Choudhury et al. 2013).[1] If another is "psychomotor retardation severe enough to be observed by others," researchers might treat slower typing speeds as a diagnostic signal (Huang et al. 2018). Systems seek out digital trace analogues for rumination and guilt in the form of a higher proportion of first-person pronoun use in Facebook posts; for feelings of worthlessness, algorithmically comparing the language we send off into the digital ether against dictionaries that group words based on their perceived emotional resonance; or for social withdrawal, lower numbers of followers and people followed and a reduced propensity to respond when tagged or at-ed. Preferably, many such signals might be combined to capture as much of the diagnostic profile as possible, drawing in physical activity, language use, vocal characteristics, and physiological data. These researchers' dream is that this sort of large-scale passive monitoring for depression systems might supplement or even replace traditional clinical diagnosis. Instead of by a clinician in a consulting room, users might be flagged by systems deployed by platforms, governments, or private companies, and perhaps without their knowledge.

What might get lost in this process of translation from psychiatric criteria to machine learning? What assumptions about depression and its sufferers get baked in to these "digital phenotypes" (Zulueta et al. 2018) and "behavioral fingerprints" (De Choudhury et al. 2013)? We know that human diagnoses are fallible—as well as racist, sexist, and classist in ways that gatekeep mental healthcare from those who need it. Will these systems incorporate those biases as well, obscuring them behind the rhetoric of precision and accuracy that surrounds AI? If diagnosis takes place continuously as part of constant, passive monitoring, do the labels we use to describe who we are begin to blur and flicker? Today, one system might tag me as depressed based on a combination of my step count, my vocal register, and the filters I use on Instagram. Tomorrow, another system declares me shipshape thanks to my phone metadata and the vocabulary of my comments on Reddit. The DSM is not unchangeable, it's true. But revisions to the text take years of debate and compromise, the upshot of which is that diagnostic identities don't change overnight. As diagnosis becomes algorithmic, though, different systems register different depressive symptoms from runtime to runtime, remaking what constitutes depression on the fly. And given that many of these tools work by leveraging passively collected social media data, we shouldn't be surprised that they carry with them a tacit presence for the kinds of "meaningful connections" platform giants claim to promote. They model a lack of depression as compulsory platform sociality, rewriting mental health as knowing how to use your phone, or an Instagram filter, or a hashtag in the way that platforms prefer.

But let's assume for a moment that these translations are all perfectly valid. We'd still be faced with a different set of problems. These systems aim to diagnose the way a clinician might, if a clinician could watch you and millions of others post, speak, work, and rest—a hyperattentive, unobtrusive clinician who

followed you every second of your life, watching over you as you scrolled and slept. But AI depression diagnostics built on the DSM would not mimic the way that clinicians actually use the DSM—or, more precisely, don't use it. For one, clinicians don't mechanically slot patients into diagnoses based on the criteria. As Luhrmann (2001, 42) recounts, they can exercise broad professional latitude, opting for a more "optimistic" diagnosis with a better prognosis (e.g., manic-depression instead of schizophrenia) or for one that reads as less stigmatizing (e.g., dysthymia instead of depression). They may also deliver gatekeeping diagnoses shaped by the biases of their profession and of society writ large: Men get alcohol use disorder while women get depression (Harrington 2019, 182); white boys get autism while Black men get schizophrenia (Metzl 2010). Psychiatrists rely more on intuition than checklists (Luhrmann 2001), and they're not beholden to the DSM's diagnostic codes (Greenberg 2013, 68). AI tools that try to automate symptom checklists imagine human clinicians as high priests of the DSM, scrying its vagaries and scrupulously applying its dicta. But nothing could be further from the truth. Instead of creating algorithmic equivalents for human clinicians, they reify a diagnostic practice that never was.

Recently, some systems use deep learning to do just that, gleaning supposed signs of depression without human supervision, directly from the data itself.

Some researchers have pointed out that this process of translation is itself a conservative approach. After all, if diagnostic categories aren't connected to a psychiatric ground truth in the first place, why try to recreate their criteria as machine learning features? Why not simply let the system learn the signals of depression from scratch? Recently, some systems use deep learning to do just that, gleaning supposed signs of depression without human supervision, directly from the data itself. Some reports seem to indicate that these more radical approaches are significantly better at identifying depression (Squires et al. 2023), an improvement that comes with serious trade-offs. As neural networks recognize a dog not from any human-like concept of "looks-like-Lassie," or "has-tail," or "lolls-tongue," instead focusing on inscrutable pixel-by-pixel gradients, so these deep learning diagnostics generate symptom categories that might be completely uninterpretable from any psychiatric theory or cultural understanding of depression.

If AI tools elevate diagnostic criteria that are mostly observed in the breach, or create new symptoms that may correspond in no way to our understandings of depression, there is another distinction that is both simpler and more radical. A clinician may give you a diagnosis, but they may just as easily not. Many are of the professional opinion that a diagnosis is largely a bureaucratic shibboleth that unlocks the coffers of insurance companies, permitting reimbursement for medicine and care (Greenberg 2013). In the absence of these institutional imperatives, why diagnose? Providing a diagnosis might also foreclose other kinds of self-understanding for a patient, so clinicians may simply withhold whatever possibilities pass through their heads, hanging diagnostic fire. AI diagnostic tools, on the other hand, classify; it's their very nature. Data goes in, "depressed" or "not-depressed" comes out. They are incapable of the most crucial activity: simply listening, simply waiting, for the next thing to be said, for something else to happen.

The New Economy of Diagnosis

Who's a diagnosis for, anyway? Traditionally, we might think of it as a service provided to the patient. This is not to sideline the kinds of forced diagnosis that take place in psychiatric emergency rooms, in prisons, and in schools. But for many of us, a diagnosis is something we seek out, a process we engage in more or less voluntarily. We, or our insurer, pay for the sequence of numbers and letters the clinician appends to our medical records. We have our reasons: A diagnosis might get us the accommodations we need at work or school, might enable us to afford psychiatric medicine or therapy. It might also provide what Catherine Tan (2018) calls "biographical illumination." Here, a diagnostic label can enable new understandings of one's own behavioral and neurological difference. It can have a reparative function, playing a constitutive role in a new, more self-accepting identity and granting access to supportive communities.

The AI tools I've described promise to rewire the political economy or ecology of diagnosis: who seeks one, who pays, who provides, and why. Instead of an individual seeking out a diagnosis, sharing personal information voluntarily with a clinician, tools that passively monitor large populations for depression transform platforms and devices into distributed consulting rooms, collecting what may be HIPAA-protected information without consent for tech companies. When they mine our social media accounts for signals of distress, they do not respect the contexts of our disclosures: maybe you felt safe enough on your favorite subreddit or a close friends–only Instagram story to wonder if there was some more constitutive cause to your bad feelings. As is common with big data surveillance, these tools don't so much destroy privacy rights as redistribute them. While we can't keep our disclosures from feeding the machine, the data sets themselves are transformed into corporate assets. And, while anyone can go out and buy a copy of the DSM-5-TR, these systems' algorithms can be locked away behind layers of technical opacity and intentional secrecy.

These tools do not ask us if we *want* to be diagnosed, and, typically, they render a diagnosis *not to us*, but to a third party: a research team, a tech startup, a government agency, an employer. At the worst, your diagnosis becomes a product, another psychographic data point to be brokered and exploited in surveillance capitalism, as Amazon patents (Brodkin 2018) and Facebook leaks (Levin 2017) reveal is already the case. But even assuming the best intentions, these diagnoses feed a coercive "asymmetric paternalism" (Schüll and Zaloom 2011), or what we might call a "neuropaternalism," singling out those deemed defective for psychiatric assessment and correction.

The tools I've described above, even the ones that remain academic proofs-of-concept, instantiate this neuropaternalism and lay the groundwork for its expansion. This infrastructure for the new political economy of diagnosis emerges precisely at the permeable boundary between university research and the mental health and technology industries. Take, for instance, Aiberry, Inc., an AI diagnosis startup premised on the idea that "current mental health practices are tedious, subjective, and error-prone." The seed for Aiberry's underlying technology grew from academic research aimed at automated depression detection from facial micromovements, later expanded to incorporate vocal and linguistic signals. The work to validate its models was undertaken by teams of university researchers at Georgetown, the University of Texas at Austin, the University of Arizona, and the University of Bradford. These models are

not, however, the kinds of prototypes that some of these systems remain. Instead, they form the guts of a corporate wellness program that serves AI assessments to employees, and returns to employers "quantified metrics showing true ROI" for their investment in algorithmic diagnostics.



Screenshot of Aiberry website.

This is the new political economy of diagnosis in action. Aiberry is not alone here. Many systems are currently or eventually intended for the use of employers and insurers, like those developed by Kintsugi and Ginger.io, the latter of which has now been folded into the corporate wellness behemoth, Headspace. Given the role that they play in shifting the political economy of diagnosis, we might well wonder what happens to those who contest them. When a tech company contracted by your employer to reduce the "burden of depression" on their bottom line warns you that you "might be at risk" and "prompts you to seek care," as Apple's Seabreeze envisions, what happens if you don't seek the kind of care the system, the company, and your employer deem appropriate?

Articles that examine the harms of new technologies often conclude with rousing calls to action. Instead of throwing all of our ever-watchful devices out the window, we can imagine some countermeasures that promise some degree of individual resistance to AI neuropaternalism. Similar to TrackMeNot (Howe and Nissenbaum 2018)—a browser extension that aims to prevent surveillance of web searches by shrouding each genuine query in a cloud of decoys—we might imagine DiagnoseMeNot, obfuscating digital phenotypes tied to depression by cloaking them in cheerful artificial noise. Or, along the lines of makeup meant to defeat facial recognition algorithms (Alvarez 2014), we could imagine filters that scrupulously overwrite our facial micromovements, accelerate our speech, and perk up our diction. But, as Os Keyes (2021) argues in the context of antisurveillance makeup, these individual interventions might function more as resistance signaling and fashion statement than effective sabotage. They may even distract from the kinds of collective resistance that could throw a wrench in the new AI economy of diagnosis.

If contemporary diagnosis and big data mining meet at the "end of theory," one meaningful first step might come from theory itself. There is a role here for those of us in disability theory who think about

disability critically, and differently from the engineers, tech companies, and employers invested in systems like Aiberry, Kintsugi, Ginger.io, and Seabreeze. Since the 1970s, depression has been one of the most thoroughly biomedicalized disabilities. Instead of Cvetkovich's "public feeling," it's often viewed from the perspective of the psy-sciences and the psychiatric industrial complex as the unfortunate but apolitical result of bad chemicals sloshing around in the brain, to be treated by other chemicals, created and marketed by pharmaceutical corporations, and administered by licensed professionals. Unlike, for instance, Deafness, depression is rarely viewed as a politicized disability identity that can serve as the basis for collective action. These new AI tools naturalize that depoliticization. They obscure that process, and the deep conceptual and practical shifts it involves, behind a rhetoric of precision and accuracy. Instead of a potential identity, they constitute depression as an "epidemic" and a "burden," and constitute all of as potentially "at risk," our place along that statistical gradient to be determined by close surveillance of every tap and text and swipe. But they also evacuate the possible resistance depression as an identity might provide, replacing a potential collective engaged in self-advocacy with atomized platform users. Until we can articulate a disability politics of depression, we might be left with this technopolitics, one that repurposes automated diagnosis as a path to maximum return-on-investment.

Recommended Readings

Cvetkovich, Ann, and Abby Wilkerson. 2016. "Disability and Depression." *Journal of Bioethical Inquiry* 13 (December): 497–503. https://doi.org/10.1007/s11673-016-9751-z.

Ehrenberg, Alain. 2010. *The Weariness of the Self: Diagnosing the History of Depression in the Contemporary Age*. Montreal: McGill-Queen's University Press.

Greenberg, Gary. 2013. *The Book of Woe: The DSM and the Unmaking of Psychiatry*. New York: Penguin.

Semel, Beth M. 2022. "Listening Like a Computer: Attentional Tensions and Mechanized Care in Psychiatric Digital Phenotyping." *Science, Technology, & Human Values* 47 (2):266–290. https://doi.org/10.1177/01622439211026371.

Zeavin, Hannah. 2021. *The Distance Cure: A History of Teletherapy*. Cambridge, MA: The MIT Press.

References

Abraham, Tara. 2016. *Rebel Genius: Warren S. McCulloch's Transdisciplinary Life in Science*. Cambridge, MA: The MIT Press.

Alvarez, Ana Cecilia. 2014. "How to Hide from Big Brother." *Dazed Digital*, March 5, 2014. https://www.dazeddigital.com/artsandculture/article/19131/1/artists-writers-show-how-to-hide-from-big-brother-government-surveillance.

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*, June 23, 2008. https://www.wired.com/2008/06/pb-theory/.

Brodkin, Jon. 2018. “Amazon Patents Alexa Tech to Tell if You’re Sick, Depressed and Sell you Meds.” *ArsTechnica*, October 11, 2018. https://arstechnica.com/gadgets/2018/10/amazon-patents-alexa-tech-to-tell-if-youre-sick-depressed-and-sell-you-meds/.

Browne, Grace. 2022. “The Problem with Mental Health Bots.” *Wired*, October 1, 2022, https://www.wired.com/story/mental-health-chatbots/.

Chancellor, Stevie, Michael L. Birnbaum, Eric D. Caine, Vincent M.B. Silenzio, and Munmun De Choudhury. 2019. “A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media.” In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 79–88. New York: Association for Computing Machinery. https://doi.org/10.1145/3287560.3287587.

Cvetkovich, Ann. 2012. *Depression: A Public Feeling*. Durham, NC: Duke University Press.

Cvetkovich, Ann, and Abby Wilkerson. 2016. “Disability and Depression.” *Journal of Bioethical Inquiry* 13 (December): 497–503. https://doi.org/10.1007/s11673-016-9751-z.

Davies, William. 2015. *The Happiness Industry: How the Government and Big Business Sold Us Well-being*. New York: Verso.

De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. “Predicting Depression via Social Media.” *Proceedings of the International AAAI Conference on Web and Social Media* 7 (1):128–137. https://doi.org/10.1609/icwsm.v7i1.14432.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter.” *PloS one* 6 (12):e26752. https://doi.org/10.1371/journal.pone.0026752.

Ehrenberg, Alain. 2010. *The Weariness of the Self: Diagnosing the History of Depression in the Contemporary Age*. Montreal: McGill-Queen’s University Press.

Eichstaedt, Johannes C., Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, and H. Andrew Schwartz. 2018. “Facebook Language Predicts Depression in Medical Records.” *Proceedings of the National Academy of Sciences* 115 (44):11203–11208. https://doi.org/10.1073/pnas.1802331115.

Garg, Parie, and Sam Glick. 2018. “AI’s Potential to Diagnose and Treat Mental Illness.” *Harvard Business Review*, October 22, 2018. https://hbr.org/2018/10/ais-potential-to-diagnose-and-treat-mental-illness.

Greenberg, Gary. 2013. *The Book of Woe: The DSM and the Unmaking of Psychiatry*. New York: Penguin.

Greene, Jeremy A. 2022. *The Doctor Who Wasn't There: Technology, History, and the Limits of Telehealth*. Chicago: The University of Chicago Press.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457 (7232):1012–1014. https://doi.org/10.1038/nature07634.

Guo, Xiaobo, Yaojia Sun, and Soroush Vosoughi. 2021. "Emotion-based Modeling of Mental Disorders on Social Media." In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 8–16. New York: Association for Computing Machinery. https://doi.org/10.1145/3486622.3493916.

Harrington, Anne. 2019. *Mind Fixers: Psychiatry's Troubled Search for the Biology of Mental Illness*. New York: WW Norton & Company.

Howe, Daniel, and Helen F. Nissenbaum. 2009. "Trackmenot: Resisting Surveillance in Web Search." In *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society*, edited by Ian Kerr, Carole Lucock, and Valerie Steeves, 417–440. Oxford: Oxford University Press.

Huang, He, Bokai Cao, Phillip S. Yu, Chang-Dong Wang, and Alex D. Leow. 2018. "dpMood: Exploiting Local and Periodic Typing Dynamics for Personalized Mood Prediction." In *2018 IEEE International Conference on Data Mining (ICDM)*, 157–166. Singapore: IEEE. https://doi.org/10.1109/ICDM.2018.00031.

Inkster, Becky, Shubhankar Sarda, and Vinod Subramanian. 2018. "An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study." *JMIR mHealth and uHealth* 6 (11):e12106. https://doi.org/10.2196/12106.

Jablonsky, Rebecca. 2022. "Meditation Apps and the Promise of Attention by Design." *Science, Technology, & Human Values* 47 (2):314–336. https://doi.org/10.1177/01622439211049276.

Keyes, Os. 2021. "Now You See It." *Real Life*, October 28, 2021. https://reallifemag.com/now-you-see-it/.

Kisliuk, Bill. 2020. "UCLA Launches Major Mental Health Study to Discover Insights about Depression." UCLA Newsroom. August 4, 2020. https://newsroom.ucla.edu/releases/ucla-launches-major-mental-health-study-to-discover-insights-about-depression.

Kramer, Adam D. I. 2010. "An Unobtrusive Behavioral Model of 'Gross National Happiness.'" In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 287–290. New York: Association for Computing Machinery. https://doi.org/10.1145/1753326.1753369.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-

scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24):8788–8790. https://doi.org/10.1073%2Fpnas.1320040111.

Kubie, Lawrence S. 1930. "A Theoretical Application to Some Neurological Problems of the Properties of Excitation Waves which Move in Closed Circuits." *Brain* 53, no. 2 (July):166–177. https://doi.org/10.1093/brain/53.2.166.

Levin, Sam. 2017. "Facebook Told Advertisers It Can Identify Teens Feeling 'Insecure' and 'Worthless'." *The Guardian*, May 1, 2017. https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens.

Luhrmann, Tanya M. 2001. *Of Two Minds: An Anthropologist Looks at American Psychiatry*. New York: Vintage.

McCulloch, Warren S. 1949. "The Brain Computing Machine." *Electrical Engineering* 68 (6):492–497. https://doi.org/10.1109/EE.1949.6444817.

McGinnis, Ellen W., Steven P. Anderau, Jessica Hruschak, Reed D. Gurchiek, Nestor L. Lopez-Duran, Kate Fitzgerald, Katherine L. Rosenblum, Maria Muzik, and Ryan S. McGinnis. 2019. "Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood." *IEEE Journal of Biomedical and Health Informatics* 23 (6): 2294–2301. https://doi.org/10.1109/JBHI.2019.2913590.

Metzl, Jonathan M. 2010. *The Protest Psychosis: How Schizophrenia Became a Black Disease*. New York: Beacon Press.

Nagy, Jeffrey Scott. 2022. "Watching Feeling: Emotional Data from Cybernetics to Social Media." PhD diss., Stanford University. https://purl.stanford.edu/qc500bj9156.

Pickering, Andrew. 2010. *The Cybernetic Brain: Sketches of Another Future*. Chicago: The University of Chicago Press.

Quilantan, Bianca. 2018. "In a Fight Against Depression, UCLA Relies on Technology." *Chronicle of Higher Education*, March 8, 2018. https://www.chronicle.com/article/in-a-fight-against-depression-ucla-relies-on-technology/.

Reece, Andrew G., and Christopher M. Danforth. 2017. "Instagram Photos Reveal Predictive Markers of Depression." *EPJ Data Science* 6 (1):15. https://doi.org/10.1140/epjds/s13688-017-0110-z.

Sadowsky, Jonathan. 2020. *The Empire of Depression: A New History*. New York: John Wiley & Sons.

Schüll, Natasha Dow, and Caitlin Zaloom. 2011. "The Shortsighted Brain: Neuroeconomics and the Governance of Choice in Time." *Social Studies of Science* 41 (4):515–538.

https://doi.org/10.1177/0306312710397689.

Squires, Matthew, Xiaohui Tao, Soman Elangovan, Raj Gururajan, Xujuan Zhou, U. Rajendra Acharya, and Yuefeng Li. 2023. “Deep Learning and Machine Learning in Psychiatry: A Survey of Current Progress in Depression Detection, Diagnosis and Treatment.” *Brain Informatics* 10 (1):1–19. https://doi.org/10.1186/s40708-023-00188-6.

Stark, Luke. 2018. “Algorithmic Psychometrics and the Scalable Subject.” *Social Studies of Science* 48 (2): 204–231. https://doi.org/10.1177/0306312718772094.

Tan, Catherine D. 2018. “‘I’m a Normal Autistic Person, Not an Abnormal Neurotypical’: Autism Spectrum Disorder Diagnosis as Biographical Illumination.” *Social Science & Medicine* 197 (January): 161–167. https://doi.org/10.1016/j.socscimed.2017.12.008.

Tsugawa, Sho, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. “Recognizing Depression from Twitter Activity.” In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3187–3196. New York: Association for Computing Machinery. https://doi.org/10.1145/2702123.2702280.

Turkle, Sherry. 2005. *The Second Self: Computers and the Human Spirit*. Cambridge, MA: The MIT Press.

UCLA. n.d. “UCLA Grand Challenges: Depression.” Accessed May 20, 2023. https://grandchallenges.ucla.edu/depression/.

Wells, Georgia, Jeff Horowitz, and Deepa Seetharaman. 2021. “Facebook Knows Instagram is Toxic for Teen Girls, Company Documents Show.” *Wall Street Journal*, September 14, 2021. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739.

Wilson, Elizabeth A. 2010. *Affect and Artificial Intelligence*. Seattle: University of Washington Press.

Winkler, Rolfe. 2021. “Apple Is Working on iPhone Features to Help Detect Depression, Cognitive Decline.” *Wall Street Journal*, September 21, 2021. https://www.wsj.com/articles/apple-wants-iphones-to-help-detect-depression-cognitive-decline-sources-say-11632216601.

Zeavin, Hannah. 2021. *The Distance Cure: A History of Teletherapy*. Cambridge, MA: The MIT Press.

Zulueta, John, Andrea Piscitello, Mladen Rasic, Rebecca Easter, Pallavi Babu, Scott A. Langenecker, Melvin McInnis, et al. 2018. “Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A Biaffect Digital Phenotyping Study.” *Journal of Medical Internet Research* 20 (7):e241. https://doi.org/10.2196/jmir.9775.

Footnotes

1 Researchers who leverage social media activity seem rarely to consider the role that platforms themselves play in sleep disturbance—a role familiar to anyone who’s found themselves doomscrolling until 3 a.m.