Artificial and Human Intelligence in Mental Health

Mariano Sigman, Diego Fernandez Slezak, Lucas Drucaroff, Sidarta Ribeiro, Facundo Carrillo

■ *While technology has dramatically changed medical practice, various aspects of mental health practice and diagnosis remain almost unchanged across decades. Here we argue that artificial intelligence — with its capacity to learn and infer from data the workings of the human mind — may rapidly change this scenario. However, this process will not happen without friction and will promote an explicit reflection of the overarching goals and foundational aspects of mental health. We suggest that the converse relation is also very likely to happen. The application of artificial intelligence to a field that relates to the foundations of what makes us human — our volition, our thoughts, our pains and pleasures — may shift artificial intelligence back to its earliest days, when it was mostly conceived of as a laboratory to explore the limits and possibilities of human intelligence.*

The Beginning of a Dream

Artificial intelligence (AI), combined with the challenges of social interactions and mental health, primarily invites us to raise our awareness of how AI influences our social interactions and our mental health; ideally, these elements combined should help us to live a healthier life. Many scholars have explored how AI is already changing the way humans relate to one another, and how this may well lead to human mental overload when machines take over all routine work, leaving to humans only the more taxing nonroutine tasks. In this article, however, we focus on the bidirectional interaction of AI and mental health — that is, on how AI allows us to chart and expand human cognition. And, with this, how AI is likely to change forever the human perception of self.

In the early twentieth century, several scholars were advocating for different approaches to psychotherapy. During this time of heated discussion, Carl Rogers developed a humanistic view, largely based on empathy — a method known as *person-centered* (now *Rogerian*) psychotherapy (Rogers 1978).

Interestingly, although empathy seems to be at the foundation of human feelings, it is gated by a very simple mechanism of imitation and mimicry, which is pervasive and automatic (Pfeifer, Iacoboni, Mazziotta, and Dapretto 2008) — and, hence, quite simple to instantiate into an algorithm. This led to a most intriguing paradox in the early sixties on the origins of AI.

Joseph Weizenbaum, at the Massachusetts Institute of Technology, programmed one of the first chat bots in history. It was deliberately simple and shallow, and it could not be otherwise, given the rudimentary computational resources available. Weizenbaum's program, named ELIZA[1] after a character in George Bernard Shaw's play *Pygmalion*, was based on very simple rules. It sought to identify keywords in the input that were mapped (as partial imitations) to stereotyped responses. It was also programmed to solve

confusing solutions with a simple but seemingly intelligent rule: When no keywords were recognized in the input, the program responded "Please, go on" or "Elaborate on this" (Weizenbaum 1966). Here is an example of a dialog (retrieved from Morice 1987):

Computer: What is your problem?
Patient: I am worried about the weather.
Comp: How long have you been worried about the weather?
Pt: Since my mother passed away.
Comp: Tell me more about your family.
Pt: They are all into computers.
Comp: Do machines frighten you?
Pt: No.
Comp: Please do not be so short with me.
Pt: I am sorry.
Comp: Earlier you talked about your mother.
Pt: Yes, she didn't like me very much.
Comp: Please go on.[2]

ELIZA's strategy, re-assembling the input as part of the response, becomes quite evident. And although now it may seem naive and simple, it also illustrates clearly how this simple trickery of imitation and mimicry can build a sense of empathy. Psychology has shown that, in all dimensions of personality, we tend to like people that resemble us (Byrne and Nelson 1965); this explains why talking to a machine that mirrors what we are saying provides a feeling of comfort. In fact, to Weizenbaum's surprise, early users of ELIZA attributed human-like feelings to it. This may have led to the idea of using it to implement DOCTOR, a version of ELIZA that simulated a Rogerian psychotherapy. The arch-humanist approach to psychotherapy, in the early 1960s, was being replaced by an IBM 7094,[3] a computer 10,000 times larger than a current cell-phone and 107 times slower. ELIZA's story was colorful and intriguing. After all, it provided a strong proof of concept that some aspects of human communication, which may seem at first extremely sophisticated, actually rely on quite simple rules. Some scholars even believed that this approach could be used in practical terms for therapy. With regard to this last claim — which apparently was never Weizenbaum's intention — ELIZA's contributions, if any, were quite modest.

A few years later, on the other coast of the USA, at Stanford University, John Colby, inspired by ELIZA's intriguing capacity to emulate human conversations developed PARRY.[4] This software was an upgraded version with an additional production system that aimed to identify the context and current state of the conversation. In other words, PARRY had a broad idea of what the topic or the intention of the communication is about (Hutchens 1996). PARRY's main distinctive aspect was that it was instantiated with a paranoid personality and had specific protocols to recognize whether it was being provoked and, if so, respond sparingly.

PARRY's testing ground went beyond qualitative judgments, to a more challenging quantitative and objective examination. A group of experienced psychiatrists interviewed both PARRY and real paranoid patients. Then, a different (blind) group received these interviews, and was unable to distinguish the conversations held with the real paranoids from the ones emulated by PARRY (Colby, Hilf, Weber, and Kraemer 1972). This experiment set a remarkable foundation for an automated psychiatry and, at the same time and not coincidentally, represented a step toward a machine capable of passing a Turing Test.[5]

The capacities and limits of PARRY, and to a lesser degree, ELIZA, have been broadly discussed elsewhere. They were undoubtedly preliminary and rudimentary steps, and hence their practical utility today is negligible. Yet, understanding how they set a path for AI to revolutionize mental health is, we believe, a fertile and instructive process of thought.

An Ethical Foundation for Mental Health

The extraordinary progression of computer science and AI over the last few years is hard to miss. If PARRY and ELIZA could even hint at a solution for an automated, objective, and algorithmic approach to psychiatry, then it is clear that current technologies provide a whole new range of solutions. Rudiments of vague and imprecise intelligent programs have become unbeatable standards over a wide range of domains, such as chess (Silver et al. 2018), stock forecasting (Oncharoen and Vateekul 2018), and radiology (Esteva et al. 2017; McKinney et al. 2020). A priori, it may seem that the same should happen soon in mental health. And, to some degree, as we revise below, this is very likely to happen. But before going into this revision it seems necessary to make explicit a fundamental difficulty for an ethical foundation of AI in mental health, namely the lack of a precise and a universally agreed-upon value function that establishes what an optimal diagnosis or treatment should seek for.

In chess intelligence, just to mention an example, while the search space and combinatorial of possible moves may be arbitrarily complex, the objective to be reached by the intelligence seems quite clear and simple: *Win as many games as possible*. Even there, one could argue about subtle parameters of the objective function. One may prioritize a program to play creatively, or to play minimizing risk, but there would not be major arguments on the settings of what this intelligence is poised to do. Setting the goals for AIs in all aspects that define us as humans or societies poses a much greater challenge. This has been amply revised, for instance, in the case of self-driving vehicles, that eventually will have to make hard moral decisions that may involve prioritizing some lives or individuals over others (Bonnefon, Shariff, and Rahwan 2016). Having to explicitly write down these decisions in a program forces us to think on aspects of morality and ethics that are often left implicit or subject to vague reasoning

(Awad et al. 2018). The advent of AI in psychiatry will be challenged by the concrete and explicit pressure to delimit objective and formal procedures: when to diagnose, how to react, and when not to overreact, do not only define a notion of mental health. These questions will have direct implications for how to deal with a conundrum: the coexistence of the great complexity in a vast plurality of minds, and the pressure to define a notion of normality. We predict that these challenges are going to be remarkably difficult and not free of friction, but they will also catalyze and promote a necessary expansion of the mental space required to define the overarching objectives, aims, and values of mental health.

In fact, it is not easy to come up today with a standardized notion of mental health. A good effort, for instance, can be found in the patient care information of the Mayo Clinic:

> Mental illness … refers to a wide range of mental health conditions — disorders that affect your mood, thinking and behavior. ... Many people have mental health concerns from time to time. But a mental health concern becomes a mental illness when ongoing signs and symptoms cause frequent stress and affect your ability to function. A mental illness can make you miserable and can cause problems in your daily life, such as at school or work or in relationships.[6]

This definition is undoubtedly a good effort to provide a compact definition of a remarkably complex problem. However, it also strongly hints that a large number of concepts and parameters remain implicit. Just to name a few: the notions that define functionality and how they relate to dynamic social expectations; how the inability to function may be relevant above and beyond suffering; the notion and ownership of autonomy and volition; the pain of an individual and the pain of others; when and where to set the thresholds that define mental health; the boundaries and asymmetries between seeking pleasure and avoiding pain… These questions have been subjected to ongoing discussion since the inception of psychiatry, but the imminent arrival of AI in mental health makes the questions explicit and certainly more urgent.

The Fuzzy Boundaries of Mental Health and Disease

In the previous section we revised how AI will put pressure to define overarching aims of a discipline that inevitably sets what we think are desirable mental experiences. Along the same line, at a finer grain, there is the issue of specific diagnosis, which relates to the existence of inner categories that underlie the organization of the mind. The writer Jorge Luis Borges exquisitely described in his essay "The Analytic Language of John Wilkins" how we humans compulsively seek to organize the universe into categories, even when we know for certain that those are inevitably temporary and imperfect. The parcellation of mental health through the definition and classification of diseases has also been a major challenge in the history of psychiatry (Insel 2014). In current clinical practice, patients are categorically diagnosed based on specific guidelines, such as the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) (American Psychiatric Association 2013). However, to no surprise, some authors are very critical of the DSM, arguing that it lacks scientific validity (Frances and Widiger 2012; Gøtzsche-Astrup and Moskowitz 2016). Part of the disagreement comes from the empirical observation that, among patients diagnosed with the same disorder, there is a considerable heterogeneity concerning etiology and pathophysiological and symptomatic expression (Allsopp et al. 2019). And, conversely, there are often multiple symptomatic similarities among patients receiving different diagnoses (Allsopp et al. 2019; Lee et al. 2019; Mitelman 2019). The mechanistic links between biology and psychiatry remain tenuous.

The US National Institute of Mental Health acknowledged the earlier problems and set out to develop a noncategorical system for mental health investigation, named Research Domain Criteria, to implement an evidence-based psychiatric classification (Insel et al. 2010). This is a dimensional approach that aims to address constructs such as emotion, cognition, motivation, and social behavior independently of the DSM diagnosis. In fact, this organization of mental health can also provide quantitative measures of mental portraits even in mentally healthy subjects (Cuthbert and Insel 2013). Instead of defining broad health categories, the focus is set to simple specific functions — for example, social communication, working memory, or reward responsiveness — to identify underlying mechanisms, causes of disruption, and potential treatments (Insel et al. 2010).

The imprecision and need for improvement in diagnosis poses both a challenge and an opportunity for AI in mental health. Seeing how AI has progressed over the last few years, it seems reasonable to assume that this interaction may progress in two steps: first simply providing (as we revise below) objective and quantitative tools that may help in the production of existing diagnosis and treatment criteria; and second, acknowledging that, by providing these tools, AI may help us redefine these boundaries so that the overarching goals of mental health become more attainable.

A Taxonomy of AIs in Mental Health

Psychologic and cognitive science has recently gone through a crisis of validity and reproducibility (Peng 2015). The reasons that led to this crisis are at the heart of what makes psychiatric diagnosis complex. The human mind is remarkably variable, with a very large and intricate number of dimensions that are mixed up in scientific studies. The problem becomes worse when studies are conducted (as has been historically) using relatively small samples.

A solution to this problem, specifically relevant for the intersection of AI and mental health, is our current capacity to amass enormous amounts of expressions of the human mind through an organized corpora of user behaviors and virtually infinite text repositories. In combination with advances in Machine Learning (ML) that provide ways to learn from these vast repositories of data, this advance offers an unprecedented opportunity for the inquiry of the human mind, both in health and in illness.

Psychiatric diagnosis has relied, since its early days, on a conversation between psychiatrist and the patient. During these interviews, the physician identifies different features including speech content and speech structure, and also a wide variety of nonverbal aspects of communication. These attributes are combined (explicitly or implicitly) to elucidate the patient's mental state, toward a diagnosis (Cowen, Harrison, and Burns 2012). Language, a privileged window into the human mind, has been at the heart of psychiatric diagnosis.

It is hard to precisely delimit the process of thought and reasoning by which a psychiatrist forms a diagnosis from speech information with the aim of instantiating it into an algorithm. The first difficulty, of course, is that this process varies very widely between different practitioners. As an attempt to somehow mitigate this source of variance, several semistructured interviews have been proposed and used in clinical applications, such as Structured Clinical Interview for DSM disorders, the Minnesota Multiphasic Personality Inventory, and the Structured Interview of Prodromal Syndromes and Scale of Prodromal Symptoms questionnaires (Lemos, Vallina, Fernandex, and Ortega 2006; Lobbestael, Leurgans, and Arntz 2011). The responses are collapsed into a single number that guides the definition of symptoms and diagnosis. The advantage of this procedure is achieving a certain degree of objectivity and quantitative assessment. This comes at the cost of restricting and conditioning the space of exploration and interaction toward a diagnosis.

But intrinsic variability is not the only, nor even the main problem, towards an algorithm that may emulate the process of a psychiatrist coming up with a diagnosis. One could focus this process on highly experienced psychiatrists, who are more likely to detect subtle symptoms more rapidly and precisely. But even then, when they are asked to explain how they arrived at certain conclusions, most of the time they recur to qualitative, semi-intuitive answers (Sadock and Sadock 2011). This difficulty is not unique to the inquiry of mental health; instead, it is ubiquitous across all cognitive research: Introspection is an opaque thing, and as a consequence we are most often unaware of the mental reasoning by which we accomplish extraordinary feats (Corallo, Sackur, Dehaene, and Sigman 2008; Marti, Sackur, Sigman, and Dehaene 2010; Shalom et al. 2013).

A chess analogy may help us understand this fundamental limitation. Classic scholars in psychology, championed by the remarkable work of de Groot (2008), have sought to understand how chess masters think and with this, more generally, the processes of human reasoning and decision-making. The same strategy was used in AI. For a long time, chess programs were fueled by grandmasters (experts) and a team of interpreters and translators that identified, from the largely distorted reports of process of thoughts, elements to build a value function that could guide the search procedure. More recently, however, as is widely known, the approach to AI in chess changed dramatically, using remarkable computer power and a great insight of doing convolutions in arbitrary deep spaces, to promote a process of self-discovery (Campbell, Joseph Hoane, and Hsu 2002; Silver et al. 2018). And, while the introspection of deep-networks is even more opaque than that of humans, and hence this approach may not seem useful to guide human thoughts, it has provided remarkable new insights that have been subsequently used in human conceptions. For instance, in the famous second game of the match of Alpha-Go against master Lee Sedol, the computer made a move that humans would have never even considered.[7] The evidence that this way of playing is effective, contrary to all our prior intuitions and understanding, subsequently changed how humans approach the game. A similar process is highly likely to develop in AI applied to mental health. At a first stage AI should be able to convert human expert reasoning into algorithms for diagnosis. At a second stage, AI will introduce novel ideas that psychiatrists may incorporate into their reasoning and diagnostic process.

One of the most relevant tools of AI for mental health is natural language processing (NLP) that serves to interpret and respond to natural human language (Jurafsky and Martin 2014). Many methods and strategies in NLP have been developed in the last few years to characterize different features of mental health (Berisha, Wang, LaCross, and Liss 2015; Voleti, Liss, and Berisha 2019). A few years ago, we applied NLP as an effort to automate and synergize psychiatric early diagnosis and prevention. We investigated the capacity of NLP to predict the development of psychosis in clinical high risk (CHR) patients (Bedi et al. 2015). Individuals identified as CHR were followed-up for 2.5 years and labeled as *converter* (CHR+) if they had a psychotic episode during this period of time, or *non-converter* (CHR–) if no psychotic episode was present. This protocol began with a long interview, where Structured Interview of Prodromal Syndromes and Scale of Prodromal Symptoms scales were measured, and an open-ended interview was performed by a specialist. Combining NLP for feature extraction from this first interview with ML methods for pattern recognition, we were able to sort with high accuracy the CHR+ from the CHR patients. We validated these methods in a second multisite study performed with a larger group of patients (Corcoran et al. 2018).

Depression has also been widely explored as a condition target for AI support. It is estimated that 6.7% of American adults suffer from this condition. Some years ago, Eric Horvitz and colleagues studied how Twitter may be used as a lens to study and prevent depression in populations (De Choudhury, Counts, and Horvitz 2013; Corcoran et al. 2018). This study has been repeated, followed-up, and extended in many other mental health conditions (see De Choudhury, Counts, Horvitz, and Huff 2014; Coppersmith, Dredze, and Harman 2014).

NLP can be used to go beyond predicting mental conditions, and help specialists decide the most effective treatments. For example, therapies using psychoactive drugs have been explored, as an alternative treatment of treatment-resistant depression (Osório, Sanches, de Macedo, and Dos Santos 2015; Palhano-Fontes et al. 2019; Sanches et al. 2016; Scott and Carhart-Harris 2019). These treatments have shown promising results but with wide variability, with some patients showing great responsiveness to the treatment and others little effect. A few years ago, we showed that NLP based on interviews of patients performed before psilocybin treatment can distinguish those patients for whom the treatment will be effective, in contrast with those for whom it will not (Carrillo et al. 2018). These results open new perspectives on how AI may synergize and catalyze the development and success of new treatments.

NLP has also been used in one of the domains in which rapid detection is of greatest urgency: suicidal risk. Suicide ideation has been widely explored using automated text analysis in electronic medical records (Fonseka, Bhat, and Kennedy 2019). The analysis of electronic medical records provided useful information about which features from text were most important to detect and estimate suicidal risk. These features were used to analyze internet documents and find records using the search utility Google to estimate suicidal risk, with better performance in some cases than with classic scales (Ma-Kellams, Or, Baek, and Kawachi 2015; Song, Song, Seo, and Jin 2016).

These examples illustrate the benefit of monitoring patients' online activity to sample data continuously, instead of sporadically doing so in psychiatric consultation. At the same time, it also clearly raises the issue and difficulty that this may convey on privacy. How public information available in social networks may or may not be used to prevent and improve mental health is still not regulated and has to be part of an important ongoing ethical debate. An insightful review of the conflict between privacy and the necessity of acquiring data to avoid risk can be found in Fonseka, Bhat, and Kennedy (2019).

This article does not intend to be an exhaustive review of how all the different AI technologies may assist mental health. It is mostly focused on NLP because analysis of language has been at the core of psychiatric diagnosis. However, this does not in any way intend to imply that other approaches are less promising or relevant. For example, a qualitatively different approach has been based on an analysis of behavior and decision-making, mostly relying on reinforcement learning and Bayesian models to quantify these analyses (Adams, Huys, and Roiser 2016). Some pathologies (Malloy-Diniz, Miranda, and Grassi-Oliveira 2017) may manifest particular symptoms that could be useful to analyze with a specifically designed behavioral task, instead of with a general analysis of language.

Detecting risk and diagnosing and understanding the etiology of a condition are only the first necessary steps to understand the best course of action to help a patient, which brings us back to the heated discussion from the times of Carl Roger and Weizenbaum's ELIZA. Interestingly, ELIZA and PARRY took an extremely early "go" at this greater challenge: in moving beyond language understanding, they produced reactions that were expected to be meaningful, productive, and empathetic. Identifying the best of a set of possible treatments falls within classic optimization problems that are remarkably well suited for ML and AI (Sutton and Barto 1998; Goodfellow, Bengio, and Courville 2016). It requires identifying a state and evaluating the consequences of distinct possible courses of actions (treatments). In a world in which data are structured and well organized, this appears to be, despite the huge dimensionality and variance, a tractable problem.

A first step toward this aim is to automatize therapy, as in modern versions of ELIZA. Initiatives with complex graphical user interfaces emulating therapists have shown promising results in cognitive treatments (Stratou et al. 2015). In some cases, paradoxically, empathic virtual agents and empathetic robots may be even more effective compared with human therapies as, for many reasons (including privacy, stress, and distraction), people may feel more comfortable with them in talking about private and sensitive issues (Costa et al. 2018). Digital therapy has also proven to be an effective procedure for treatment of substance abuse (Waltz 2018).

AI is likely to change almost every aspect of mental health in the near future: assisting and quantifying diagnosis; defining the diagnostic categories and boundaries; changing the process by which patients communicate with physicians; dramatically modifying the rate, quality, and form by which practitioners can monitor and follow patient evolution; and identifying for each patient (with genetic, cultural, and temporal variability) the combination of therapies that might be most useful in a patient-centered practice of medicine.

A Human AI

In their quest to conceive a thinking machine, the founders of modern computers circumstantially developed a program to understand human intelligence. For the conception of his device, Turing

observed and sought to emulate his own thoughts (Zylberberg, Dehaene, Roelfsema, and Sigman 2011). Then, and for many years, AI was mostly driven by the practical necessity of solving a myriad of complex problems, leading to solutions that did not resemble, were not inspired by, and could not be pertinent to, the study of human intelligence. But with the development of deep convolutional networks (Sermanet and LeCun 2011; Kalchbrenner, Grefenstette, and Blunsom 2014; Tygert, Bruna, Chintala, and LeCun 2016; Jia et al. 2018), we have witnessed a remarkable explosion of this program and in this process, several scholars in computer science and AI have raised again the idea that inspiration in human intelligence (the best known example of how to grow a mind) can both fuel and give new directions to AI (Lake, Ullman, Tenenbaum, and Gershman 2016). Maybe AI has an opportunity to go back to Turing's early intention and become a laboratory to simulate ourselves, to explore the limits and possibilities of the human mind. By changing priors and combination rules in image recognition algorithms, people have begun to ask about the transition from perception to dreams, revisiting Philip K. Dick's famous question: *Do Androids Dream of Electric Sheep?*[8]

Here we have discussed how the deeply human roots of mental health will pose challenges to AI, and how this will oblige our society to think explicitly about its foundational aspects: the boundaries of health and disease, categories of diagnosis, and objectives of treatment. The converse relation, we think, is also very likely to manifest. The pressure of applying AI to the conceptions of what makes us human — the secrets of our volition, our thoughts, our ideas, our pains, and our pleasure — will steer AI back to its inception: an exploration of the limits and possibilities of human intelligence, desires, and dreams.

Notes

1. psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm

2. Cited from Morice, R. 1987. Artificial Intelligence and Psychiatry. *American Journal of Psychiatry* 69(21): 1352–3.

3. www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP7094.html

4. www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/classics/parry/0.html

5. Alan M. Turing's Turing Test, or the "imitation game," introduced by him in 1950. plato.stanford.edu/entries/turing-test/

6. www.mayoclinic.org/diseases-conditions/mental-illness/symptoms-causes/syc-20374968

7. www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol

8. Dick, Philip K. 1968. Do Androids Dream of Electric Sheep? Phoenix, AZ: Orion.

References

Adams, R. A., Huys, Q. J. M. and Roiser, J. P. 2016. 'Computational Psychiatry: towards a mathematically informed understanding of mental illness', *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(1), pp. 53–63.

Allsopp, K. et al. 2019. 'Heterogeneity in psychiatric diagnostic classification', *Psychiatry Research*, 279, pp. 15–22.

American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.; and Rahwan, I. 2018. The Moral Machine Experiment. *Nature* 563(7729): 59–64. doi.org/10.1038/s41586-018-0637-6

Bedi, G.; Carrillo, F.; Cecchi, G.; Fernández Slezak, D.; Sigman, M.; Mota, N.; Ribeiro, S.; Javitt, D.; and Copell, M. 2015. Automated Analysis of Free Speech Predicts Psychosis Onset in High-Risk Youths. *NPJ Schizophrenia* 1(1): 15030. doi.org/10.1038/npjschz.2015.30

Berisha, V.; Wang, S.; LaCross, A.; and Liss, J. 2015. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. *Journal of Alzheimer's Disease* 45(3): 959–63. doi.org/10.3233/JAD-142763

Bonnefon, J.-F.; Shariff, A.; and Rahwan, I. 2016. The Social Dilemma of Autonomous Vehicles. *Science* 352(6293): 1573–6. doi.org/10.1126/science.aaf2654

Borges, Jorge Luis. 1952. "The Analytical Language of John Wilkins", Other Inquisitions (1937–1952). Sur.

Byrne, D., and Nelson, D. 1965. Attraction as a Linear Function of Proportion of Positive Reinforcements. *Journal of Personality and Social Psychology* 1(6): 659–63. doi.org/10.1037/h0022073

Campbell, M.; Joseph Hoane, A.; and Hsu, F.-H. 2002. Deep Blue. *Artificial Intelligence* 134(1–2): 57–83. doi.org/10.1016/S0004-3702(01)00129-1

Carrillo, F.; Sigman, M.; Fernandez Slezak, D.; Ashton, P.; Fitzgerald, L.; Stroud, J.; Nutt, D.; and Carhart-Harris, R. L. 2018. Natural Speech Algorithm Applied to Baseline Interview Data Can Predict which Patients Will Respond to Psilocybin for Treatment-Resistant Depression. *Journal of Affective Disorders* 230(1): 84–6. doi.org/10.1016/j.jad.2018.01.006

Colby, K. M.; Hilf, F. D.; Weber, S.; and Kraemer, H. C. 1972. Turing-Like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes. *Artificial Intelligence* 3: 199–221. doi.org/10.1016/0004-3702(72)90049-5

Coppersmith, G.; Dredze, M.; and Harman, C. 2014. Quantifying Mental Health Signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Stroudsburg, PA: Association for Computational Linguistics. doi.org/10.3115/v1/W14-3207

Corallo, G.; Sackur, J.; Dehaene, S.; and Sigman, M. 2008. Limits on Introspection: Distorted Subjective Time during the Dual-task Bottleneck. *Psychological Science* 19(11): 1110–7. doi.org/10.1111/j.1467-9280.2008.02211.x

Corcoran, C. M.; Carrillo, F.; Fernández-Slezak, D.; Bedi, G.; Klim, C.; Javitt, D. C.; Bearden, C. E.; and Cecchi, G. A. 2018. Prediction of Psychosis Across Protocols and Risk Cohorts Using Automated Language Analysis. *World Psychiatry* 17(1): 67–75. doi.org/10.1002/wps.20491

Costa, A. P.; Charpiot, L.; Rodríguez Lera, F.; Ziafati, P.; Nazarikhorram, A.; van der Torre, L.; and Steffgen, G. 2018. More Attention and Less Repetitive and Stereotyped Behaviors Using a Robot with Children with Autism. In *2018 27th Institute of Electrical and Electronics Engineers*

(IEEE) International Symposium on Robot and Human Interactive Communication (RO-MAN). Piscataway, NJ: IEEE. doi.org/10.1109/ROMAN.2018.8525747

Cowen, P.; Harrison, P.; and Burns, T. 2012. *Shorter Oxford Textbook of Psychiatry*. Oxford, UK: Oxford University Press. doi.org/10.1093/med/9780199605613.001.0001

Cuthbert, B. N., and Insel, T. R. 2013. Toward the Future of Psychiatric Diagnosis: The Seven Pillars of RDoC. *BMC Medicine* 11(1): 126. doi.org/10.1186/1741-7015-11-126

De Choudhury, M.; Counts, S.; and Horvitz, E. 2013. Social Media as a Measurement Tool of Depression in Populations. In *Proceedings of the 5th Annual Association for Computing Machinery (ACM) Web Science Conference on Web Science*. New York: ACM. doi.org/10.1145/2464464.2464480

De Choudhury, M.; Counts, S.; Horvitz, E.; and Huff, A. 2014. Characterizing and Predicting Postpartum Depression from Shared Facebook Data. In *Proceedings of the 17th Association for Computing Machinery (ACM) Conference on Computer Supported Cooperative Work and Social Computing*. New York: ACM. doi.org/10.1145/2531602.2531675

de Groot, A. D. 2008. *Thought and Choice in Chess*. Amsterdam, The Netherlands: Amsterdam University Press. www.aup.nl/en/book/9789053569986/thought-and-choice-in-chess

Esteva, A.; Kuprel, B.; Novoa, R. A.; Ko, J.; Swetter, S. M.; Blau, H. M.; and Thrun, S. 2017. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* 542(7639): 115–8. doi.org/10.1038/nature21056

Fonseka, T. M.; Bhat, V.; and Kennedy, S. H. 2019. The Utility of Artificial Intelligence in Suicide Risk Prediction and the Management of Suicidal Behaviors. *The Australian and New Zealand Journal of Psychiatry* 53(10): 954–64. doi.org/10.1177/0004867419864428

Frances, A. J., and Widiger, T. 2012. Psychiatric Diagnosis: Lessons From the DSM-IV Past and Cautions for the DSM-5 Future. *Annual Review of Clinical Psychology* 8(1): 109–30. doi.org/10.1146/annurev-clinpsy-032511-143102

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. Cambridge, MA: The MIT Press

Gøtzsche-Astrup, O., and Moskowitz, A. 2016. Personality Disorders and the *DSM-5*: Scientific and Extra-Scientific Factors in the Maintenance of the Status Quo. *The Australian and New Zealand Journal of Psychiatry* 50(2): 119–27. doi.org/10.1177/0004867415595872

Hutchens, J. L. 1996. How to Pass the Turing Test by Cheating. School of Electrical, Electronic and Computer Engineering Research Report TR97-05. Perth, Australia: University of Western Australia.

Insel, T. R. 2014. The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry. *The American Journal of Psychiatry* 171(4): 395–7 doi.org/10.1176/appi.ajp.2014.14020138

Insel, T. R.; Cuthbert, B.; Garvey, M.; Heinssen, R.; Pine, D. S.; Quinn, K.; Sanislow, C.; and Wang, P. 2010. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *The American Journal of Psychiatry* 167(7): 748–51. doi.org/10.1176/appi.ajp.2010.09091379

Jia, Y. H.; Bai, L.; Wang, P.; Guo, J. L.; and Xie, Y. X. 2018. Deep Convolutional Neural Network for Correlating Images and Sentences. In *International Conference on Multimedia Modeling*, 154–65. Lecture Notes in Computer Science, vol. 10704. Berlin, Germany: Springer. doi.org/10.1007/978-3-319-73603-7_13

Jurafsky, D., and Martin, J. H. 2014. *Speech and Language Processing*. London, UK: Pearson.

Kalchbrenner, N.; Grefenstette, E.; and Blunsom, P. 2014. A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 655–65. Stroudsburg, PA: Association for Computational Linguistics. doi.org/10.3115/v1/P14-1062

Lake, B. M.; Ullman, T. D.; Tenenbaum, J. B.; and Gershman, S. J. 2016. Building Machines that Learn and Think Like People. *Behavioral and Brain Sciences* 40: e253. doi.org/10.1017/S0140525X16001837

Lee, P. H.; Anttila, V.; Won, H.; Feng, Y. C. A.; Rosenthal, J.; Zhu, Z.; Tucker-Drob, E. M.; Nivard, M. G.; Grotzinger, A. D.; Posthuma, D.; and Wang, M. M. 2019. Genome Wide Meta-Analysis Identifies Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *BioRxiv* 528117. doi.org/10.1101/528117

Lemos, S.; Vallina, O.; Fernandex, P.; and Ortega, J. A. 2006. Predictive Validity of the Scale of Prodromal Symptoms (SOPS). *Actas Españolas de Psiquiatría* 34(4): 216–23.

Lobbestael, J.; Leurgans, M.; and Arntz, A. 2011. Inter-Rater Reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clinical Psychology & Psychotherapy* 18(1): 75–9. doi.org/10.1002/cpp.693

Ma-Kellams, C.; Or, F.; Baek, J. H.; and Kawachi, I. 2015. Rethinking Suicide Surveillance: Google Search Data and Self-Reported Suicidality Differentially Estimate Completed Suicide Risk. *Clinical Psychological Science* 4(3): 480–4. doi.org/10.1177/2167702615593475

Malloy-Diniz, L. F.; Miranda, D. M.; and Grassi-Oliveira, R. 2017. Editorial: Executive Functions in Psychiatric Disorders. *Frontiers in Psychology* 8: 1461. doi.org/10.3389/fpsyg.2017.01461

Marti, S.; Sackur, M.; Sigman, M.; and Dehaene, S. 2010. Mapping Introspection's Blind Spot: Reconstruction of Dual-Task Phenomenology Using Quantified Introspection. *Cognition* 115(2): 303–13. doi.org/10.1016/j.cognition.2010.01.003

McKinney, S.; Sieniek, M.; Godbole, V.; Godwin, J.; Antropova, N.; Ashrafian, H.; Back, T.; Chesus, M.; Corrado, G.; Darzi, A.; Etemadi, M.; Garcia-Vicente, F.; Gilbert, F. J.; Halling-Brown, M.; Hassabis, D.; Jansen, S.; Karthikesalingam, A.; Kelly, C. J.; King, D.; Ledsam, J. R.; Melnick, D.; Mostofi, H.; Peng, L.; Reicher, J. J.; Romera-Paredes, B.; Sidebottom, R.; Suleyman, M.; Tse, D.; Young, K. C.; De Fauw, J.; and Shetty, S. 2020. International Evaluation of an AI System for Breast Cancer Screening. *Nature* 577(7788): 89–94. doi.org/10.1038/s41586-019-1799-6

Mitelman, S. A. 2019. Transdiagnostic Neuroimaging in Psychiatry: A Review. *Psychiatry Research* 277(July): 23–38. doi.org/10.1016/j.psychres.2019.01.026

Morice, R. 1987. Artificial Intelligence and Psychiatry. *The American Journal of Psychiatry* 144(10): 1352–3. doi.org/10.1176/ajp.144.10.1352

Oncharoen, P., and Vateekul, P. 2018. *Deep Learning for Stock Market Prediction Using Event Embedding and Technical Indicators*. New York: Institute of Electrical and Electronics Engineers. doi.org/10.1109/ICAICTA.2018.8541310

Osório, F. d. L.; Sanches, R. F.; Macedo, L. R.; dos Santos, R. G.; Maia-de-Oliveira, J. P.; Wichert-Ana, L.; de Araujo, D. B.; Riba, J.; Crippa, J. A.; and Hallak, J. E. 2015. Antidepressant Effects of a Single Dose of Ayahuasca in Patients

with Recurrent Depression: A Preliminary Report. *Revista Brasileira de Psiquiatria (Sao Paulo, Brazil)* 37(1): 13–20. doi.org/10.1590/1516-4446-2014-1496

Palhano-Fontes, F.; Barreto, D.; Onias, H.; Andrade, K. C.; Novaes, M. M.; Pessoa, J.; Mota-Rolim, S. A.; Osório, F. d. L.; Sanches, R.; dos Santos, R. G.; Tófoli, L. F.; de Oliveira Silveira, G.; Yonamine, M.; Riba, J.; Santos, F. R.; Silva-Junior, A. A.; Alchieri, J. C.; Galvão-Coelho, N. L.; Lobão-Soares, B.; Hallak, J. E. C.; Arcoverde, E.; Maia-de-Oliveira, J. P.; and Araújo, D. B. 2018. Rapid Antidepressant Effects of the Psychedelic Ayahuasca in Treatment-Resistant Depression: A Randomized Placebo-Controlled Trial. *Psychological Medicine* 49(4): 655–63. doi.org/10.1017/S0033291718001356

Peng, R. 2015. The Reproducibility Crisis in Science: A Statistical Counterattack. *Significance* 12(June 15): 30–2. doi.org/10.1111/j.1740-9713.2015.00827.x

Pfeifer, J. H.; Iacoboni, J.; Mazziotta, J. C.; and Daprettoa, M. 2008. Mirroring Others' Emotions Relates to Empathy and Interpersonal Competence in Children. *NeuroImage* 39(4): 2076–85. doi.org/10.1016/j.neuroimage.2007.10.032

Rogers, C. R. 1978. Carl Rogers on Personal Power. Arrow Books.

Sadock, B. J., and Sadock, V. A. 2011. Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry. Lippincott Williams & Wilkin.

Sanches, R. F.; de Lima Osório, F.; dos Santos, R. G.; Macedo, R. H.; Maia-de-Oliveira, J. P.; Wichert-Ana, L.; de Araujo, D. B.; Riba, J.; Crippa, J. A. S.; and Hallak, J. E. C. 2016. Antidepressant Effects of a Single Dose of Ayahuasca in Patients with Recurrent Depression. *Journal of Clinical Psychopharmacology* 36(1): 77–81. doi.org/10.1097/JCP.0000000000000436

Scott, G., and Carhart-Harris, R. L. 2019. Psychedelics as a Treatment for Disorders of Consciousness. *Neuroscience of Consciousness* 2019(1): niz003. doi.org/10.1093/nc/niz003

Sermanet, P., and LeCun, Y. 2011. Traffic Sign Recognition with Multi-Scale Convolutional Networks. In *The 2011 International Joint Conference on Neural Networks*. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE). doi.org/10.1109/IJCNN.2011.6033589

Shalom, D. E.; de Sousa Serro, M. G.; Giaconia, M.; Martinez, L. M.; Rieznik, A.; and Sigman, M. 2013. Choosing in Freedom or Forced to Choose? Introspective Blindness to Psychological Forcing in Stage-Magic. *PLoS One* 8(3): e58254. doi.org/10.1371/journal.pone.0058254

Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Matthew, L.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; Lillicrap, T.; Simonyan, K.; and Hassabis, D. 2018. A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go Through Self-Play. *Science* 362(6419): 1140–4. doi.org/10.1126/science.aar6404

Song, J.; Song, T. M.; Seo, D.-C.; and Jin, J. H. 2016. Data Mining of Web-Based Documents on Social Networking Sites that Included Suicide-Related Words among Korean Adolescents. *The Journal of Adolescent Health* 59(6): 668–73. doi.org/10.1016/j.jadohealth.2016.07.025

Stratou, G.; Morency, L.-P.; DeVault, D.; Hartholt, A.; Fast, E.; Lhommet, M.; Lucas, G.; Morbini, F.; Georgila, K.; Scherer, S.; Gratch, J.; Marsella, S.; Traum, D.; and Rizzo, A. 2015. A Demonstration of the Perception System in SimSensei, a Virtual Human Application for Healthcare Interviews. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. Piscataway, NJ: IEEE. doi.org/10.1109/ACII.2015.7344661

Sutton, R. S., and Barto, A. G. 1998. Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks* 9(5): 1054. doi.org/10.1109/TNN.1998.712192

Tygert, M.; Bruna, J.; Chintala, S.; and LeCun, Y. 2016. A Mathematical Motivation for Complex-Valued Convolutional Networks. *Neural Computation* 28(5): 815–25. doi.org/10.1162/NECO_a_00824

Voleti, R. N. U.; Liss, J.; and Berisha, V. 2019. A Review of Automated Speech and Language Features for Assessment of Cognition and Thought Disorders. *IEEE Journal of Selected Topics in Signal Processing* 14(2): 282–98. doi.org/10.1109/JSTSP.2019.2952087

Waltz, E. 2018. Pear Approval Signals FDA Readiness for Digital Treatments. *Nature Biotechnology* 36(6): 481–2. doi.org/10.1038/nbt0618-481

Weizenbaum, J. 1966. ELIZA: A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM* 9(January): 36–45. doi.org/10.1145/365153.365168

Zylberberg, A.; Dehaene, S.; Roelfsema, P. R.; and Sigman, M. 2011. The Human Turing Machine: A Neural Framework for Mental Programs. *Trends in Cognitive Sciences* 15(7): 293–300. doi.org/10.1016/j.tics.2011.05.007

Mariano Sigman is a researcher at the Neuroscience Lab, Universidad Torcuato Di Tella; and at the School of Languages and Education, Nebrija University. He has experience in neuroscience and education, decision-making, political corruption, networks, graphs, the inner workings of the human brain, computational psychiatry, and the language of thought.

Diego Fernandez Slezak is a full-time professor at the Computer Science Department, Universidad de Buenos Aires. He is head of the Applied Artificial Intelligence Lab. His research focuses on using state-of-the-art ML models and developing novel ML models in health applications, such as NLP and medical imaging.

Lucas Drucaroff is a clinical psychiatrist (MD) and researcher (PhD), and a post-doc at the National Scientific and Technical Research Council (Argentina). He is a specialist in statistics for health sciences and a university teacher at the Universidad de Buenos Aires. He also works at the Instituto FLENI-National Scientific and Technical Research Council (Argentina). Drucaroff's research focuses on neuropsychological aspects of neuropsychiatric patients in relation to brain activity, including ML functional medical resonance imaging methods. In recent years, he has been working with natural language models in psychiatry.

Sidarta Ribeiro is a full professor of neuroscience and a vice-director of the Brain Institute at the Universidade Federal do Rio Grande do Norte. He has experience in neuroethology, molecular neurobiology, and systems neurophysiology, with an interest in memory, sleep, and dreams, neuronal plasticity, vocal communication, symbolic competence in nonhuman animals, computational psychiatry, and neuroeducation, as well as psychedelics and drug policy.

Facundo Carrillo is a full-time researcher in computer science at the Universidad de Buenos Aires. He works at the Applied Artificial Intelligence Lab, National Scientific and Technical Research Council (Argentina). Carrillo's research focuses on developing new natural language models with applications to computational neuroscience.