



Clinical Review Article

Deep Learning and Geriatric Mental Health

Howard Aizenstein, M.D., P.h.D., Raeanne C. Moore, P.h.D., Ipsit Vahia, M.D., Adam Ciarleglio, P.h.D., M.S.

ARTICLE INFO

Article history:

Received September, 28 2023

Revised November, 24 2023

Accepted November, 27 2023

Keywords:

Deep learning
geriatric mental health
machine learning
neuroimaging

ABSTRACT

The goal of this overview is to help clinicians develop basic proficiency with the terminology of deep learning and understand its fundamentals and early applications. We describe what machine learning and deep learning represent and explain the underlying data science principles. We also review current promising applications and identify ethical issues that bear consideration. Deep Learning is a new type of machine learning that is remarkably good at finding patterns in data, and in some cases generating realistic new data. We provide insights into how deep learning works and discuss its relevance to geriatric psychiatry. (Am J Geriatr Psychiatry 2024; 32:270–279)

Editorial accompaniment, please see page XXX.

Highlights

- **What is the primary question addressed by this study?**
What is the relevance of deep learning to geriatric psychiatry?
- **What is the main finding of this study?**
Deep learning provides new insights into geriatric psychiatry along with ethical concerns.
- **What is the meaning of the finding?**
Deep learning has broad relevance to geriatric psychiatry.

From the Department of Psychiatry (HA), University of Pittsburgh School of Medicine, Pittsburgh, PA; Department of Psychiatry (RCM), University of California San Diego, San Diego, CA; Division of Geriatric Psychiatry (IV), Harvard Medical School, Boston, MA; and the Department of Biostatistics and Bioinformatics (AC), George Washington University, Washington, D.C. Send correspondence and reprint requests to Howard Aizenstein, M.D., P.h.D., 3501 Forbes Ave, Rm 520.11, Pittsburgh, PA 15213. e-mail: aizen@pitt.edu

© 2023 The Authors. Published by Elsevier Inc. on behalf of American Association for Geriatric Psychiatry. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

<https://doi.org/10.1016/j.jagp.2023.11.008>

INTRODUCTION

Geriatric psychiatry is a study of human and medical complexity. Mental health in late-life represents the cumulative outcome of a number of factors that include changes to the brain (both age-related and pathological), changes in the body, evolving social circumstances, and psychological factors (both protective and detrimental). It can be impacted by biological determinants present at birth (e.g., genetic risks) and highly situational circumstances (e.g., bereavement). Moreover, there is growing recognition that treatment outcomes can be quite heterogeneous, many factors can impact medication response and successful psychotherapy requires tailoring. The process of providing care in this domain is thus immersed in complexity.

When imagined through the lens of data science, geriatric mental health care represents a process of gathering and incorporating a diverse set of streams of data (e.g., behavioral assessments, cognitive evaluation, medical examination, psychosocial context, imaging, genetic, passively sensed, and laboratory tests) and making both situational and longer-term decisions. With the ever-growing availability of tools to support clinical decision-making, the field of geriatric mental health should be seen as a fertile ground for incorporating data processing and predictive analytics into clinical workflow. There are many ways in which advances in computational science may help manage this complexity - from processing brain images to providing in-the-moment access to medical knowledge. The challenge is to develop proficiency in the use of these increasingly sophisticated tools while prioritizing the human aspect of care that is a foundational element of care in older adults.¹ It is increasingly clear that the future of medical science will require proficiency in the use of a new generation of artificial intelligence (AI)-based tools, but in order for these tools to achieve their potential, they should be used to augment, rather than replace, the human element of care. A relatively clear way in which this can be accomplished is by using these tools to help clinicians simplify the complexity inherent in late-life mental health care.

Our intent in this review is to provide an overview and perspective on the clinical use of an especially compelling branch of AI – deep learning. We are currently experiencing a “deep learning revolution.”² In this

revolution, computers have become so good at finding patterns in big data that solutions are being found for questions we did not know we had. For instance, using the popular smartphone app made by iNaturalist (https://www.inaturalist.org/pages/seek_app) one can now point their phone at a plant or animal to identify the species. However, multiple critical questions remain unanswered, including where the most useful applications of these tools may be, what steps need to be taken to ensure ethical development and use of these tools (including the data used to create them), and what regulatory frameworks may be necessary. The process of answering these questions will call for unprecedented dialogue and collaboration between fields and experts who may not typically collaborate – including patients, clinicians, data scientists, data visualization and design experts, ethicists, and regulators. A field like geriatric mental health which is inherently proficient at working with an interdisciplinary perspective is thus ideally placed to lead this process.

The goal of this overview is to help clinicians develop basic proficiency with the terminology of deep learning, and understand its fundamentals and early applications. We first describe what machine learning and deep learning represent. Next, we discuss how and why considerations of sample size differ between machine learning and deep learning. Next, we present sections that aim to familiarize readers with foundational mathematical principles that underlie deep learning and touch on how deep learning represents a new tool to help evolve our understanding of complex philosophical principles such as linguistics and theory of mind. Finally, we review how early clinical applications are being determined based on deep learning.

At the very outset, it is critical to acknowledge that in an emerging field like deep learning that relies on gathering vast amounts of data, there are certain to be complex ethical questions. The full scope of these challenges is unlikely to be understood at this time and a comprehensive discussion of the ethics of deep learning and AI is beyond the scope of this review. We discuss this further in the future directions section

MACHINE LEARNING

Machine learning is a component of the umbrella term AI. Artificial intelligence is the overall quest of

having computers act intelligently.³ Machine learning algorithms are those AI techniques that focus on the intelligent task of learning. In this case, machine learning refers to the process of having a computer find a pattern from seeing positive and negative examples of that pattern. The most common form of machine learning is referred to as supervised learning, because the examples involved include both the features (e.g., clinical, imaging, or demographic characteristics) under consideration for predicting an outcome of interest and the values for that outcome variable. Features can be unprocessed raw observed data, such as the medical image or recorded audio, or it can be fully processed and extracted variables. Examples of processed features that can be used for learning include regional atrophy measures extracted from the MRI or summary behavior rating scores, such as the Montgomery Asberg Depression Rating Scale. These features, paired with the observed outcomes, are used to guide the learning process.

Supervised learning typically falls into one of two categories: regression, when the outcome is numerical (e.g., depression score), or classification, when the outcome is a class label or group membership (e.g., type of dementia).⁴ The machine learning program uses these labeled examples to train a model (i.e., a type of prediction algorithm) that can be used to predict outcomes for new observations (observations not previously seen by the algorithm). This basic model is used for finding patterns in, for example, images, in patient clinical data, and from internet searches that are useful in predicting outcomes of interest. For instance, learning to predict the age of the patient based on their brain MRI, from seeing many examples of brain MRIs labeled with the age of each individual. In traditional machine learning, the algorithm finds a concise expression of the features, such as a decision tree, to describe how to classify new examples.

More and more, unsupervised and semi-supervised machine learning techniques are also being utilized. In unsupervised learning, algorithms learn patterns from unlabeled data.⁵ That is, all the variables used in the analysis are used as inputs, thus, the techniques are suitable for creating the labels in the data. For example, unsupervised learning techniques might be used to classify individuals diagnosed with a particular disease into previously unknown subtype classes based on different observed features. Semi-supervised models use a hybrid approach.

TRADITIONAL VERSUS DEEP LEARNING

Distinctions between traditional and deep learning continue to be debated. However, foundational to deep learning is the concept of the artificial neural network (ANN). ANNs are machine learning models that are designed to emulate the human brain and are characterized by one or more layers of interconnected nodes (neurons) that generate nonlinear representations of the input features which are useful for supervised, unsupervised, or semisupervised problems. Deep learning simply corresponds to those ANNs with multiple layers of interconnected nodes. In the context of supervised learning, fitting ANN models with more layers (i.e., deeper models) has the potential to lead to better prediction accuracy and this is one of the main reasons that deep learning has recently received considerable attention.

Deep learning differs from more traditional machine learning (e.g., logistic regression, random forests, and support vector machines) in several ways. As noted above, deep learning can be described by the complexity of the learner (i.e., depth of the ANN). Therefore, one way in which deep learning models can be distinguished from traditional machine learning models is by their need for a lot of examples.

Deep learning can also be distinguished from more traditional machine learning approaches by how it does feature selection. Traditional machine learning typically uses already preprocessed, or selected features. This can require manual selection of features (i.e., feature engineering). This entails experts identifying and extracting meaningful features for the algorithm to learn. Deep learning, when it works, is able to leverage the power of ANNs to automatically learn features and hierarchies of features from the raw data. This property of generating new data-driven features can be contrasted with manual feature engineering that is used more commonly in traditional machine learning. New advances using deep learning include Generative Adversarial Networks,⁶ which play a key role in deepfakes,⁷ and multihead attention⁸ (the transformer), which underlies large language models like ChatGPT.⁹

To illustrate the difference between traditional machine learning and deep learning, consider an analogy in which both methods attempt to learn the concept of "dog." Traditional machine learning, like

classification and regression trees, can be likened to learning about dogs through a description of their attributes. In this scenario, the algorithm is provided with a list of features such as "has fur," "domesticated," and "mammal." The learner then uses these features to construct a rule. The left panel of Figure 1 shows a classification tree as an example of a traditional machine learning method. This process involves identifying a pattern within the features that can distinguish which settings for the features defines being a dog. The result is a model capable of classifying an object as a dog based on the presence or absence of these predefined features. While deep learning models can be developed that also use this same list of features to develop a model for classifying an observation as a dog or not a dog, one of the major strengths of deep learning is that such derived feature lists are not necessary to develop a well-performing classifier. (Fig. 1)

Rather than using a list of derived features, deep learning can also take as its input, for example, images of dogs and non-dogs and learn features that can be used to distinguish dogs from non-dogs. Unlike traditional machine learning, where the learner is explicitly provided with features, in this case characteristics of the animal, the deep learning

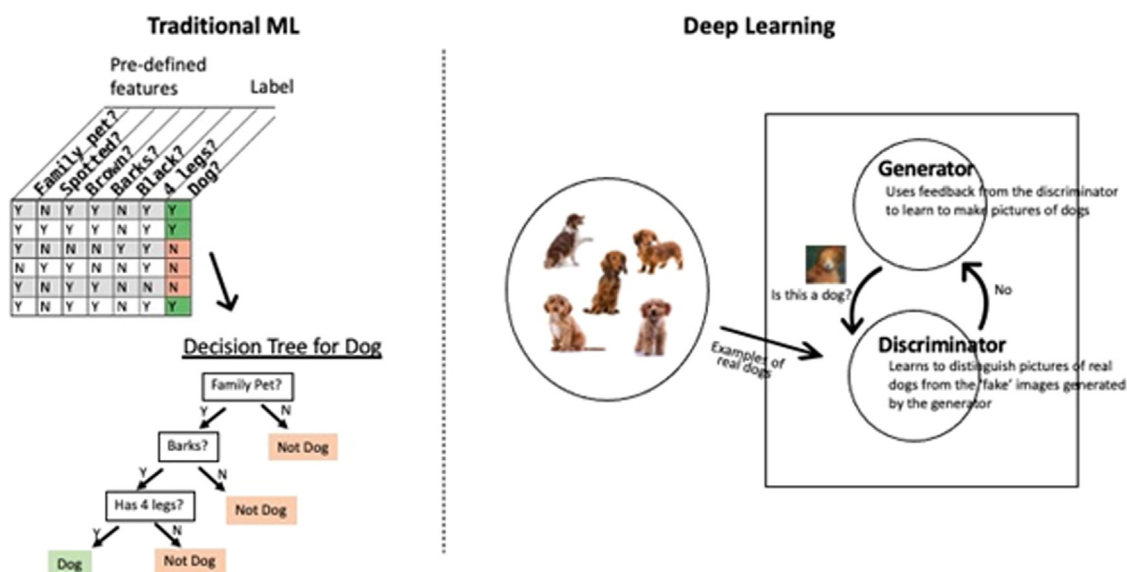
approach can extract features from the raw image data. It uses the images, rather than the extracted features. This is achieved through the use of deep (highly parameterized and flexible) networks of artificial neuron-like units. The connection strengths (i.e., weights, which are akin to slopes in regression models) between the units are adjusted to make the network improve its ability to correctly classify an image as a dog or non-dog from each labeled example.

A major challenge in the historical development of artificial neural networks was coming up with a computationally efficient way to modify the weights. The development of back-propagation, helped to make efficient estimation of deep learning models possible and this has led to widespread use of these models in many application areas. For a full description, we refer reader to Chapter 8 of an early classic book in AI and artificial neural networks.¹⁰

SAMPLE SIZE CONSIDERATIONS FOR DEEP LEARNING ALGORITHMS

As noted above, one of the very attractive features of deep learning models is that they are highly flexible. That is, they are able to capture more detail (or

FIGURE 1. Traditional machine learning versus deep learning.



nuances) of a concept, like the detail in a realistic face. This flexibility is due to the extremely large number parameters, potentially numbering in the millions, that define them. In clinical scenarios, fitting such models typically requires very large sample sizes. However, in geriatric psychiatry research, large sample sizes are just not that common – owing to a number of factors (e.g., the complexity of collecting data from geriatric patients, the time-consuming effort needed for preprocessing neuroimaging data, etc.).

An exciting recent development has been the use of “generative” deep learning methods to handle the problem of small sample sizes. This class of deep learning methods augments (artificially increases) the available training examples by synthesizing new examples from the distribution of the observed examples – thus increasing the effective sample size on which the deep learning model is trained.¹¹ A way of understanding the difference between generative AI and traditional ML is that generative AI is capable of creating new data, whereas traditional ML is limited to identifying patterns and making predictions from existing data only. In essence, generative AI leverages large amounts of data, which we have on a small number of people, to learn with fewer examples. This is exemplified in the deep learning approach of few-shot learning,¹² which shows how learning can occur with very small sample sizes. In certain ways, this can be seen as analogous to how a child is able to learn very deeply what a dog is even from knowing just one dog very well. Similarly, in medical school one can learn anatomy very well from one cadaver. One challenge with generative AI is that it can lead to certain biases in the generated sample. A recent paper showed how large language models, such as ChatGPT, may become biased towards normal appearing data, and lose representation of uncommon events that occur in the tails of the distributions.¹³

Because generative AI is creating new examples, it can be seen to learn deep associations that we would not expect. With images it can create realistic faces of no one in particular. In large language models, it may make up a reference for a study that does not exist. The AI gives the closest answer it can come up with for the question, which might be a fabricated reference.¹⁴ The fabrications are sometimes referred to as a type of AI hallucination. These fabricated results in generative AI are expected and reflect the depth of

the model. In the appropriate context, the fabrications are very helpful. However, if used naively, generative AI can easily be misleading.

DEEP LEARNING THROUGH THE LENS OF COMPUTING THEORY

With deep learning still a very nascent entity, particularly as it applies to geriatric mental health, there is an opportunity for clinicians and researchers to understand the mathematical foundations of the field. This may support a more sophisticated understanding of the explosion of research in this space that is inevitable. Here, we present a very broad overview of philosophical and historical questions that underlie the evolution of AI. One way to understand machine learning is through the lens of computing theory.¹⁵ These principles describe the inherent limits of information processing. They are used in understanding and developing machine learning approaches. One way to appreciate the contributions of the theory of computing is through its impact on cryptography (i.e., the mathematics of data security).¹⁶ The ability to create secure encryption (i.e., recoding in a new language) relies on defining the number of steps needed in an algorithm. If allowed to simply try all examples, deciphering passwords, would be easy. But, the more complex the password, the exponentially higher the number of attempts needed. This number can quickly become unimaginable, even for simple encryption problems. Thus, certain computational problems (like breaking cryptographic codes) are infeasible. However, other tasks such as identifying common patterns, and parsing certain languages can be done efficiently with a circumscribed number of steps.

The math behind computing theory relies on the assumption that the set of possible algorithms is theoretically countable, as opposed to an infinity that is so large it cannot be counted. In this case, countable means that we can explicitly describe the steps of the algorithm (sometimes referred to as the “machine”). As long as we can describe these steps, we can classify and organize them. This assumption has also been applied to neuroscience and is the basis for models of how circuits within the brain may operate. For a much deeper dive into these principles, we direct readers to the influential 1979 Pulitzer Prize-winning

TABLE 1. Traditional (PAC) Learning versus Deep (Query Learning)

Supervised Machine Learning (PAC Model)	Deep Learning (Query Learning Model)
<ul style="list-style-type: none"> • Dependent on features (x) defined by the user for predication of the response (y). • Find hidden function f, from labeled examples $f(x)=y$ that do well on new x 	<ul style="list-style-type: none"> • Allows for deep signature with New features derived from raw data (e.g., pixels of images) • May use “generative” models to • Defeat distribution limits.

book on the subject, *Godel, Escher, Bach: an Eternal Golden Braid* written by Hofstadter.¹⁷

The application of the computing theory principles to learning¹⁸ is referred to as Computational Learning Theory (COLT).¹⁹ In COLT, the generic problem of “learning” is the focus, and the properties of learning are determined by proving related theorems. Supervised machine learning is defined using parameters to circumscribe performance. An effective learning algorithm should *probably* find an *approximate* rule. This is referred to as the probably approximately correct (PAC) formalization of supervised learning.²⁰ It assumes a passive sampling of the distribution. Generative AI, on the other hand, allows for the asking of questions and an active sampling of the distribution (Table 1). Thus generative AI adds query learning to the standard example-based learning framework.²¹

In general, it is not currently possible to determine the sample size needed to obtain a well-performing deep-learning predictive model in the same way that one can compute the required sample size for estimating and testing effects under a hypothesis testing framework. Application of deep learning or any machine learning methods in geriatric psychiatry will likely require the learning of complex relationships between features derived from multiple modalities (e.g., behavioral, environmental, genetic, imaging). Having the “right” data to begin with is obviously extremely important. If one is not measuring features that are relevant for predicting responses, then neither a large sample size nor an optimal deep learning or machine learning algorithm will lead to predictions that are of any clinical relevance. We direct the reader to reviews of deep learning methods and their applications in psychiatry, with detailed discussions about sample size considerations by Koppe 2019 and Koppe 2021.^{22,23} (Table 1)

UNDERSTANDING REPRESENTATION IN DATA AND IDENTIFYING MEANING IN DATA OUTPUT

Deep learning can sometimes be approached as a black box, where one focuses only on the input and output and ignores what happens in between. A field within deep learning is representational learning, which highlights the structure of the internal representation. This is similar to the latent space or principal components that are used to determine model fit in traditional statistics. However, in deep learning, the model uses hundreds of millions of parameters to form the fit. This means that deep learning is capable of finding nonlinear combinations of features and the “deep learning” representation space can therefore exhibit properties seen in complex dynamic systems (like nuclear physics). This can facilitate aspects of deep understanding, perhaps related to meaning and even empathy. In deep learning, when the input features are recoded in particularly complex ways, it is referred to as grokking.²⁴ The term comes from the book *Stranger in a Strange Land*²⁵ by Robert Heinlein where it refers to deep empathy. Deep learning may therefore be capable as language that allows exploration of questions of meaning and end of life, often central to geriatric mental health.²⁶

EVOLVING “THEORY OF MIND” AND LINGUISTICS IN THE CONTEXT OF DEEP LEARNING

Societal understanding of computer processing has had profound impacts on our current “theory of mind.” People understand themselves with computer analogies, e.g., short and long-term memory like a computer.²⁷ Deep learning provides a more creative and flexible framework than the traditional computer model and can profoundly change “theory of mind.”

The theory of computing, as discussed above, provides a language for discussing psychological theories of mind. Prior theories have touched on related concepts of multiple-agent learning or a broader sense of distributed locus of control. Prior uses include the theory of multiple intelligences²⁸ and society of mind.²⁹ Marks-Tarlow³⁰ uses the related mathematical language of fractals as a framework for understanding

transpersonal psychology. In studying how rules for information processing are followed by computers, our brains, and our minds, we can create transdisciplinary approaches where neuroscience and computational research can inform each other.²⁷

There are, however, crucial differences. For example, Noam Chomsky, the father of computational linguistics recently described the distinction between ChatGPT and human language. Chomsky and colleagues describe that these large language models do not model the current state the way people do. Rather they focus on superficial associations. This is a key distinction between human and artificial intelligence. Chomsky demonstrated that languages and machines are essentially identical. A simple Chomsky normal form grammar is equivalent to a push down automata.¹⁵ The mathematical proofs highlight the use of a small number of “hidden” states as a canonical framework for representing language. ChatGPT and other large language models use a different approach that is more brute force for modeling language. They learn by estimating superficial associations between many features, rather than using a hidden state-based model. The current AI is based on what Chomsky describes as a superficial model of causality; therefore he does not see it as competing with human intelligence.

TECHNOLOGY TO CAPTURE RELEVANT DATA FOR DEEP LEARNING MODELS

Passive sensing: Passive sensing refers to the collection of data from various sensors embedded in everyday devices, such as smartphones and wearables, without requiring any active involvement from the user. It allows for digital phenotyping.³¹ Smartphones, in particular, can be used to continuously collect a wide range of passive sensing data, including accelerometers, gyroscopes, GPS, ambient light sensors, microphones, typing patterns, and more. Using deep learning with passive sensing data allows for the analysis of the “big data” coming from passive monitors, and for the development of potentially meaningful insights from these passive sensing technologies. Deep learning is particularly well-suited for analyzing passive sensing data due to its ability to automatically learn hierarchical representations from complex and high-dimensional data. Deep learning

models can effectively capture patterns, relationships, and dependencies within this data, enabling accurate predictions and insightful analysis. For example, Apple has integrated these tools into its devices to help with health promotion.

Deep learning applied to passive sensing data has numerous practical applications in geriatric mental health research, such as gait analysis and fall detection, emotion detection, suicide prediction, monitoring and tracking cognitive changes, health monitoring, sleep analysis, social engagement, and much more. Furthermore, deep learning approaches have been applied to sensor data with the ability to map motion. This has included a range of sensors, including infrared motion sensors, sensors on doors, sleeping mat sensors, wearable actigraphs, cameras, and radio-wave based sensors.^{32–36} A growing body of literature demonstrates how this approach has proven effective for a number of precise clinical applications in dementia care. This ranges from detecting activated behaviors such as agitation, passive behaviors such as apathy, tracking the therapeutic impact of medications such as antidepressants and antipsychotics, and monitoring the side effects of these medications. Applications also include the detection of falls and fall risk based on gait analysis. In the broader domain of neurodegenerative disorders, applying deep learning to motion detection has been demonstrated as a marker for early detection and diagnosis of Parkinson’s disease.³⁷ This approach has significant potential implications for developing early behavioral markers for Alzheimer’s disease, since behavioral impairments can manifest years before cognitive impairment.³⁸

Researchers have applied deep learning to typing patterns on a smartphone keyboard (not *what* you type, but *how* you type it) in several different neuropsychiatric populations and extracted distinctive keystroke dynamics. Further, these patterns have been found to identify unique typing signatures – or digital biomarkers – of cognition and mood. In one study that evaluated over 86,000 typing actions from 147 users (Veset et al.), keyboard dynamics data demonstrated that more severe depression was related to more variable typing speed, shorter session duration, and lower typing accuracy. Typing dynamics data has also been shown to predict future changes in mood.^{39,40}

There are several research groups examining natural language processing (NLP) and automated speech analysis in mild cognitive impairment and Alzheimer's disease to see if NLP can be used as a biomarker for Alzheimer's risk. It is also a potential marker for late-life depression.^{41–44} An NLP method was developed that recognized both behavioral activation and depressive symptoms in numerous texts exchanged among patients and therapists; behavioral activation may serve as a mediator of change in depression in patients receiving psychotherapies.⁴⁵ GPS mobility data has been identified as a digital biomarker for negative symptoms in schizophrenia,^{46,47} and there are several other examples. Ultimately, the insights derived from passive sensing data can enable personalized health interventions and real-time behavior monitoring systems.

NEUROIMAGING

There is growing literature on deep learning in neuroimaging.⁴⁸ It is being used to identify specific imaging patterns predictive of diagnosis or treatment response. Deep learning algorithms leverage the large within subject data to synthesize new data that have similar patterns. A growing open-science community specific for machine learning in AI (<https://monai.io>) makes many of these tools more accessible to the growing research community. In geriatric mental health, this has been particularly effective in identifying accurate models of the aging brain,⁴⁹ dementia,⁵⁰ and response to treatment.⁵¹

FUTURE DIRECTIONS

Deep learning is a rapidly growing field and has much potential to impact geriatric mental health beyond what we reviewed here. There are other somatic treatment approaches we did not discuss (e.g., neurostimulation) as well as additional behavioral interventions. For instance, there is the opportunity for integration of deep learning with Ecological Momentary Assessments (EMA). ChatGPT (<https://chat.openai.com/chat>) and other large language models are based on the same DL approach discussed here. ChatGPT can provide very useful guidance on a wide range of topics.

However, it uses generative AI, which involves synthetic data, and thus it can also make up things that are not true.

Overall, deep learning provides many opportunities and challenges for geriatric mental health care. NIH has identified deep learning as a target research area and describes a new era of data science (<https://www.hhs.gov/about/strategic-plan/index.html>). Industry (e.g., Meta, Amazon, Apple, Alphabet, Netflix) are currently leading research in this area. Thus, academic-industry collaborations will become essential. For many reasons there is growing recognition and concern about the ethics of developing deep learning. There are significant privacy issues in collecting the data, as well as issues around improving the informed consent process to ensure participants are appropriately informed and able to understand what data is being captured, where their data is stored, who has access, and if their data will eventually be sold. A government regulatory framework is being worked out.⁵² There are also emerging ethical concerns in the use of AI.⁵³ The way deep learning can make things up with generative AI allows for propagation of misinformation. Continued discussions will benefit from a wider understanding of deep learning and its implications.

As reviewed here, geriatric mental health can be understood using insights from deep learning. Software using deep learning offers options for mental health treatment and prevention. As we state in the Introduction, the potential for ethical challenges and bias in deep learning approaches must be considered in parallel with its potential. The risks are many and include inequitable representation of populations in data used for training models, inadequate protections of data privacy, the possibility that repurposing existing data for developing deep learning algorithms exceeds the scope of the informed consent for the original gathering of data, the potential for loss of confidentiality, and the potential for deep learning approaches being used for profitability.⁵⁴ Many of these risks have been identified in domains of medicine such as radiology and oncology where the application of deep learning is more developed. The complexities of behavioral health combined with cognitive impairment frequently seen in late life raises the potential for even more complex ethical questions. These must be acknowledged transparently and

addressed directly to maintain trust in these tools as they evolve.

AUTHOR CONTRIBUTIONS

All authors contributed to the entire manuscript.

DATA STATEMENT

This data has been presented previously by Dr. Aizenstein at the 2023 Annual Meeting of The American College of Psychiatrists on February 25, 2023, in Tucson, Arizona. The title of this talk was "Learning From Images: Computational and Neuroimaging Studies of Geriatric Mental Health."

DISCLOSURES

Raeanne Moore received salary support from NIA R01AG070956. Adam Ciarleglio received salary support from NIH K01MH113850. Howard Aizenstein received salary support from NIMH R01MH076079 and NIA RF1AG022516. Ipsit Vahia received salary support from NIA grants [P30AG073107](#), R01AG066670-S1, and R01AG066670 and the Once Upon a Time Foundation. He serves as a consultant for Otsuka and has uncompensated research collaborations with Emerald Innovations and Mirah Inc. He receives an editorial honorarium from the American Journal of Geriatric Psychiatry. None of these relationships represents a direct conflict of interest with the work presented in this manuscript.

References

1. Reynolds 3rd CF, Jeste DV, Sachdev PS, et al: Mental health care for older adults: recent advances and new directions in clinical practice and research. *World Psychiatry* 2022; 21(3):336-363; doi:[10.1002/wps.20996](#)
2. Sejnowski TJ: *The Deep Learning Revolution*. Cambridge, MA: The MIT Press, 2018:342
3. Nilsson NJ: *Principles of Artificial Intelligence*. Los Altos, Calif: Morgan Kaufmann Publishers, 1986 476 p
4. Kotsiantis S: Supervised machine learning: a review of classification techniques. *Informatica (Slovenia)* 2007; 31:249-268
5. Hofmann T: Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 2001; 42:177-196;doi:[10.1023/A:1007617005950](#)
6. Zhu J-Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017
7. Masood M, Nawaz M, Malik KM, et al: Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Appl Intell* 2023; 53(4):3974-4026;doi:[10.1007/s10489-022-03766-z](#)
8. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. 2017
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al: Large language models in medicine. *Nat Med* 2023; 29(8):1930-1940; doi:[10.1038/s41591-023-02448-8](#)
10. Rumelhart DE, McClelland JL: University of California San Diego. PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Computational Models of Cognition and Perception. Cambridge, Mass: MIT Press, 1986
11. Bowles C, Chen L, Guerrero R, et al. GAN augmentation: augmenting training data using generative adversarial networks. 2018.
12. Wang Y, Yao Q, Kwok JT, et al: Generalizing from a few examples: a survey on few-shot learning. *ACM Comput Surv*. 2020; 53(3);doi:[10.1145/3386252](#), Article 63
13. Shumailov I, Shumaylov Z, Zhao Y, et al. The curse of recursion: training on generated data makes models forget. 2023
14. McGowan A, Gui Y, Dobbs M, et al: ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. *Psychiatry Res* 2023; 326:115334;doi:[10.1016/j.psychres.2023.115334](#)
15. Hopcroft JE, Ullman JD: *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley Series in Computer Science. Reading, Mass: Addison-Wesley, 1979:418
16. Goldreich O: *Foundations of Cryptography: Volume 1: Basic Tools*. vol 1. Cambridge, UK: Cambridge University Press, 2001
17. Hofstadter DR: Gödel, Escher, Bach: An Eternal Golden Braid. New York, NY: Basic Books, 1979:777
18. Shalev-Shwartz S, Ben-David S: *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, UK: Cambridge University Press, 2014
19. Pitt L: Introduction: special issue on computational learning theory. *Machine Learning* 1990; 5:117-120
20. Valiant LG: A theory of the learnable. *Commun ACM* 1984; 27(11):1134-1142;doi:[10.1145/1968.1972](#)
21. Angluin D: Queries and concept learning. *Machine Learn* 1988; 2(4):319-342;doi:[10.1023/A:1022821128753](#)
22. Durstewitz D, Koppe G, Meyer-Lindenberg A: Deep neural networks in psychiatry. *Mole Psychiatry* 2019; 24(11):1583-1598; doi:[10.1038/s41380-019-0365-9](#)
23. Koppe G, Meyer-Lindenberg A, Durstewitz D: Deep learning for small and big data in psychiatry. *Neuropsychopharmacology* 2021; 46(1):176-190;doi:[10.1038/s41386-020-0767-z](#)
24. Liu Z, Kitouni O, Nolte NS, et al: Towards understanding grokking: an effective theory of representation learning. *Adv Neural Informat Proc Syst* 2022; 35:34651-34663
25. Heinlein RA: *Stranger in a Strange Land*. Ace trade edition. New York: Ace Books, 1991 [1991]©1991
26. Yalom ID: *Existential Psychotherapy*. New York, NY: Basic Books, 1980:524
27. Reynolds CF, 3rd, Weissman MM. Transdisciplinary science and research training in psychiatry: a robust approach to innovation. (2168-6238 (Electronic))
28. Gardner H: *Frames of Mind: The Theory of Multiple Intelligences*. New York, NY: Basic Books, 2011:467
29. Minsky M: *The Society of Mind*. New York, NY: Simon and Schuster, 1986:339 p

30. Marks-Tarlow T: A fractal epistemology for transpersonal psychology. *Int J Transpers Stud* 2020; 39(1):55-71;doi:[10.24972/ijts.2020.39.1-2.55](https://doi.org/10.24972/ijts.2020.39.1-2.55)
31. Insel TR: Digital phenotyping: a global tool for psychiatry. *World Psychiatry* 2018; 17(3):276
32. Adeli V, Korhani N, Sabo A, et al: Ambient monitoring of gait and machine learning models for dynamic and short-term falls risk assessment in people with dementia. *IEEE J Biomed Health Inform* 2023; 27(7):3599-3609;doi:[10.1109/JBHI.2023.3267039](https://doi.org/10.1109/JBHI.2023.3267039)
33. Au-Yeung WM, Miller L, Beattie Z, et al: Monitoring behaviors of patients with late-stage dementia using passive environmental sensing approaches: a case series. *Am J Geriatr Psychiatry* 2022; 30(1):1-11;doi:[10.1016/j.jagp.2021.04.008](https://doi.org/10.1016/j.jagp.2021.04.008)
34. Haslam-Larmer L, Shum L, Chu CH, et al: Real-time location systems technology in the care of older adults with cognitive impairment living in residential care: a scoping review. *Front Psychiatry* 2022; 13:1038008;doi:[10.3389/fpsy.2022.1038008](https://doi.org/10.3389/fpsy.2022.1038008)
35. Vahia IV, Kabelac Z, Hsu CY, et al: Radio signal sensing and signal processing to monitor behavioral symptoms in dementia: a case study. *Am J Geriatr Psychiatry* 2020; 28(8):820-825;doi:[10.1016/j.jagp.2020.02.012](https://doi.org/10.1016/j.jagp.2020.02.012)
36. Zhang G, Vahia IV, Liu Y, et al: Contactless in-home monitoring of the long-term respiratory and behavioral phenotypes in older adults with COVID-19: a case series. *Front Psychiatry* 2021; 12:754169;doi:[10.3389/fpsy.2021.754169](https://doi.org/10.3389/fpsy.2021.754169)
37. Yang Y, Yuan Y, Zhang G, et al: Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med* 2022; 28(10):2207-2215;doi:[10.1038/s41591-022-01932-x](https://doi.org/10.1038/s41591-022-01932-x)
38. Creese B, Ismail Z: Mild behavioral impairment: measurement and clinical correlates of a novel marker of preclinical Alzheimer's disease. *Alzheimers Res Ther* 2022; 14(1):2;doi:[10.1186/s13195-021-00949-7](https://doi.org/10.1186/s13195-021-00949-7)
39. Stange JA-O, Zulueta J, Langenecker SA, et al: Let your fingers do the talking: passive typing instability predicts future mood outcomes. (1399-5618 (Electronic))
40. Bennett CA-O, Ross MA-O, Baek E, et al: Smartphone accelerometer data as a proxy for clinical data in modeling of bipolar disorder symptom trajectory. (2398-6352 (Electronic))
41. DeSouza DD, Robin J, Gumus M, et al: Natural language processing as an emerging tool to detect late-life depression. *Front Psychiatry* 2021; 12:719125;doi:[10.3389/fpsy.2021.719125](https://doi.org/10.3389/fpsy.2021.719125)
42. Kent DM, Leung LY, Zhou Y, et al: Association of incidentally discovered covert cerebrovascular disease identified using natural language processing and future dementia. *J Am Heart Assoc* 2023; 12(1):e027672;doi:[10.1161/JAHA.122.027672](https://doi.org/10.1161/JAHA.122.027672)
43. Noori A, Magdamo C, Liu X, et al: Development and evaluation of a natural language processing annotation tool to facilitate phenotyping of cognitive status in electronic health records: diagnostic study. *J Med Internet Res* 2022; 24(8):e40384;doi:[10.2196/40384](https://doi.org/10.2196/40384)
44. Yeung A, Iaboni A, Rochon E, et al: Correlating natural language processing and automated speech analysis with clinician assessment to quantify speech-language changes in mild cognitive impairment and Alzheimer's dementia. *Alzheimers Res Ther* 2021; 13(1):109;doi:[10.1186/s13195-021-00848-x](https://doi.org/10.1186/s13195-021-00848-x)
45. Burkhardt HA, Alexopoulos GS, Pullmann MD, et al: Behavioral activation and depression symptomatology: longitudinal assessment of linguistic indicators in text-based therapy sessions. *J Med Internet Res* 2021; 23(7):e28244;doi:[10.2196/28244](https://doi.org/10.2196/28244)
46. Depp CA, Bashem J, Moore RC, et al: GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. (2398-6352 (Electronic))
47. Parrish EM, Depp CA, Moore RC, et al: Emotional determinants of life-space through GPS and ecological momentary assessment in schizophrenia: what gets people out of the house? (1573-2509 (Electronic))
48. Yan W, Qu G, Hu W, et al: Deep learning in neuroimaging: promises and challenges. *IEEE Signal Proc Magazine* 2022; 39(2):87-98;doi:[10.1109/MSP.2021.3128348](https://doi.org/10.1109/MSP.2021.3128348)
49. Cole JH, Poudel RPK, Tsagkrasoulis D, et al: Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* 2017; 163:115-124;doi:[10.1016/j.neuroimage.2017.07.059](https://doi.org/10.1016/j.neuroimage.2017.07.059)
50. Li H, Habes M, Wolk DA, et al: A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's Dement* 2019; 15(8):1059-1070;doi:[10.1016/j.jalz.2019.02.007](https://doi.org/10.1016/j.jalz.2019.02.007)
51. Squarcina L, Villa FM, Nobile M, et al: Deep learning for the prediction of treatment response in depression. *J Affect Disord* 2021; 281:618-622;doi:[10.1016/j.jad.2020.11.104](https://doi.org/10.1016/j.jad.2020.11.104)
52. Candelon F, di Carlo RC, De Bondt M, et al: *AI Regulation is Coming. How to Prepare for the Inevitable*. Brighton, Mass: Harvard Business Review, 2021
53. Sand M, Durán JM, Jongsma KR: Responsibility beyond design: Physicians' requirements for ethical medical AI. *Bioethics* 2022; 36(2):162-169;doi:[10.1111/bioe.12887](https://doi.org/10.1111/bioe.12887)
54. Prabhu SP: Ethical challenges of machine learning and deep learning algorithms. *Lancet Oncol* 2019; 20(5):621-622;doi:[10.1016/S1470-2045\(19\)30230-X](https://doi.org/10.1016/S1470-2045(19)30230-X)