Journal of Medicine, Surgery, and Public Health 1 (2023) 100026

Contents lists available at ScienceDirect

Journal of Medicine, Surgery, and Public Health

journal homepage: www.sciencedirect.com/journal/journal-of-medicine-surgery-and-public-health

Large language modeling and classical AI methods for the future of healthcare

Sri Banerjee [a,*], Pat Dunn [a,b], Scott Conard [c], Roger Ng [d]

[a] Walden University, College of Health Sciences and Public Policy, S Washington Ave, Minneapolis, MN 55401, United States
[b] American Heart Association, Center for Health Technology & Innovation, 7272 Greenville Ave, Dallas, TX 75231, United States
[c] Converging Health, 12810 Hillcrest Rd, Suite B223, Dallas, TX 75230, United States
[d] NOBELMIND, 3641 Mt. Diablo Blvd #373, Lafayette, CA 94549, United States

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Healthcare
Public health
Medicine
Precision medicine

ABSTRACT

Large Language Modeling (LLM) is ubiquitous in the healthcare industry guiding clinical decisions. With the increase in demand, we must proceed with caution in the AI industry. In this study, we evaluated the accuracy of the Random Forest model in comparison to other similar models. From the 2005 to 2010 National Health and Nutrition Examination Survey (NHANES) dataset, we assessed if there was a relationship between depression and hypertension and if depression predicted hypertension. Depression was determined using the Patient Health Questionnaire (PHQ)−9 $\geq$ 10. Hypertension was determined by taking the average of three systolic pressure readings that were elevated. Current smoking was determined by self-reported data. We tested several Random Forest models, compared with logistic regression, naïve Bayes, decision tree model and assessed these for accuracy. The percentage of the population with diabetes was 7.7%. We found that in comparison to logistic regression (87.8%), naïve Bayes (84.6%), and decision tree model (89.3%), the Random Forest model (98.4%) was considered most accurate. We also found that out of all the variables, according to the Gini impurity index, employment (150) received the highest score in relative importance. The next highest score was depression (140). This system demonstrates the importance of using traditional AI systems such as Random Forest modeling in conjunction with LLM. ChatGPT and LLM's must be further understood to integrate with classical machine learning techniques to make further advances in healthcare. LLM's have been mobilized to write history and physical assessment, extracting drug names from medical notes, and condensing radiology reports. Abstraction of medical records and other applications in healthcare can further be enhanced by using the full potential for AI systems such LLM.

Introduction

The future is being revolutionized and evolving with the advent of Large Language Modeling. Chat Generative Pre-trained Transformer (ChatGPT), a chatbot created by OpenAI, has become accessible to the general public—opening the door to many possibilities that have traditionally been considered unfathomable [1,2]. The growth of this arguably disruptive technology has been unprecedented and has been one of the fastest growing internet applications in history. Out of all the types of machine learning, ChatGPT interacts with the user in a similar way to human conversation and, on the surface, has been touted to outperform all machine learning techniques [3]. A multi-disciplinary approach can inform the public about the positive applications of Large Language Modeling (LLM) and some applications that require further investigation. ChatGPT has seen rapid growth and widespread use. In fact, ChatGPT exceeded 100 million users within a couple of months of the November 2022 release [3]. While vast swaths of the general public and researchers only know about AI through the chatbot ChatGPT, other individuals do not quite grasp the complexity of this powerful AI system [1,2]. To complicate matters, due to the popularity of this technology, there is a lot of misinformation that can be found on the internet.

With the advent of LLM with a user-friendly interface, ChatGPT has become a common business and household word. As LLM evolves, such

* Corresponding author.
E-mail addresses: srikanta.banerjee2@mail.waldenu.edu (S. Banerjee), pat.dunn@heart.org (P. Dunn), scott.conard@convergenthealth.com (S. Conard), roger.ng@sutterhealth.org (R. Ng).

https://doi.org/10.1016/j.glmedi.2023.100026
Received 30 October 2023; Received in revised form 31 October 2023; Accepted 31 October 2023
Available online 2 November 2023

comparisons will become increasingly crucial in understanding the strengths and applications of different models in this rapidly advancing field [4,5]. The models GPT-4 (similar features as ChatGPT—but many times faster and cost the company $4.6 million), Google Bard and Meena (trained on human to human interactions from public-domain social media), LLaMA (Large Language Model Meta AI), Flan-UL2 (Google Research), and BLOOM (BigScience Large Open-science Open-access Multilingual Language Model) vary significantly in their number of parameters, training data, training objectives, special features, and accessibility. Take for instance, BigScience Workshop (BLOOM's parent company), uses 176 billion parameters for the LLM to understand the context. This extensive parameter range enables BLOOM to discern intricate linguistic patterns, resulting in precise and relevant text generation. There are many more that have not been included.

Another perspective is understanding LLM from the perspective of neural networks or deep learning [6,7]. Due to the billions of parameters, LLM can process many types of information all at once. The parameters in a LLM are similar to the weights in a standard neural network. In both LLMs and neural networks, these parameters are numerical values that start as random coefficients and are adjusted during training to minimize loss. This allows for the simultaneous measurement of more conditions.

Some academicians are quick to criticize ChatGPT as a way to potentially use this technology to complete writing assignments, exclusively through plagiarism. However, in essence, there are many deep learning applications in generative AI. For example, sentiment analysis can provide quick insight about the nature of the writing as "positive", "negative", or "neutral." During the height of the COVID-19 pandemic, previously, we conducted an analysis of the patient family visitation instructions for hospitals and how unwelcoming (negative sentiment) some of the instructions were.

Instructors quickly condemn this as a place where students can plagiarize using ChatGPT without having to conduct their own research. while not acknowledging what all of the positive applications that this new technology has made possible. In the academic world, a universal alarm bell rings passing judgment—not giving this technology another thought. On the contrary, through a multidisciplinary approach, this technology should be embraced.

In order to protect and help vulnerable populations, LLM should be applied to social determinants of health (SDOH). Collecting data on the SDOH requires a thorough understanding of how medical records can be extracted and abstracted for information. For instance, understanding the connection of healthcare access and disease processes can allow improved screening questions and text prediction that are based on training data from previous responses. Even though many patients are screened for social determinants of health, having a data abstraction method through LLM can allow for a more complete picture of the patient population within a certain geographic area. This allows for informed screening questions rather than standardized scales.

The full potential is only understood when the complex interactions between social determinants of health and the disease states include LLM applications for many disorders such as acute coronary syndrome, opioid use disorder, epilepsy, and asthma [6]. There have been other applications in e-health and even plastic surgery. One example is where LLMs can be used for clinical note summarization, clinical entity recognition, and extracting ICD-10-CM Codes. These are the capabilities of the healthcare-specific large language models by John Snow labs and have a high success rate at 76% [7].

One area that is expanding in healthcare is natural language processing to "read" the narrative of the chart and abstract pertinent information. In order to detect medical errors and locate discrepancies between diagnosis and treatment, one can ensure a more accurate clinical picture [7,8]. Another example is to give LLM a PubMed abstract and ask about what the key results were gleaned. This way healthcare workers do not have to waste time unnecessarily having to read the whole article. Using random forest, researchers and trained professionals can create classifiers according to the training information provided from LLM processed chart abstraction. Similarly, information about housing access can be collected from medical records and appropriate resources can be provided to improve all the domains in social determinants of health [9,10]. However, the question that remains is that in a clinical setting, is the complete information being accurately presented when the chart information is being summarized by an LLM. After all, much of the information that will be used to treat patients is built from training information that has already been provided, which may be incomplete.

Large Language Models are one such supervised learning model is the random forest models. Random Forests and Neural Networks are different in that Neural Networks usually improve from large amounts of data and continuously improve accuracy, while Random Forests often have little performance gain at certain levels of data. Both Random Forest and LLM are considered machine learning. However, this is where the similarities end. Random forest is considered traditional machine learning, whereas LLM is considered exclusively deep learning. However, Random Forest is used as a classifier, and is of optimal value in the field of healthcare. Large Learning Models are types of deep learning such as neural networks.

As LLM technology expands tremendously, the world moves with caution before completely accepting the possibilities. This model combines ensemble learning methods with the decision tree classifier. However, it is not fully understood how Large Language Models may enhance existing machine learning models such as random forest, if at all. In this research study, we create Random Forest models and compare this to existing LLMs in order to build improved predictive models in healthcare.

Methods

A random forest classifier can be considered an integrated classification algorithm and comprised of individual decision trees. To optimize predictive accuracy, randomization of predictor variables along with bootstrap aggregation was utilized. The 2005–2010 nationally representative National Health and Nutrition Examination Survey (NHANES) was administered to all noninstitutionalized individuals and conducted by the National Center for Health Statistics. The systolic blood pressure was calculated by a mean of three measurements of $\geq$ 130 mm Hg. Depression was defined by the Patient Health Questionnaire-9. This is a nine-item screening instrument that asked the frequency of depressive symptoms in the past two weeks. A cutpoint of $\geq$ 10 was considered positive for clinically relevant depression. Other covariates include healthcare diagnosed patient reported positive response to diabetes, by asking participants "Have you ever been told by a doctor or health professional that you have _." Gender, sleep trouble, and employment were considered important covariates. Sleep trouble was determined by the question "Have you ever told a doctor or other health professional that you have trouble sleeping?" (yes/no). For those individuals that said yes, the person was considered to have trouble sleeping.

From individual level NHANES data a random forest model was built. The random forest model was split into 80% training model and 20% testing model. The train-test split is a model validation procedure that reveals how the model performs on new data. The input was mtry = 3 since the default in the statistical software is the square root of the total number of predictors for classification problems. The accuracy was predicted by comparing the performance of the random forest, binary logistic regression, decision tree model, and the naïve Bayes model. When there was a split in a node, the Gini impurity was calculated in the descendant node. The calculated Gini impurity was less than the parent node. Each time a variable had a node split, the Gini impurity criterion was calculated for the two descendent nodes. The descendent node is less than the parent node. The sum of Gini decreases for each variable across all the trees in the forest. R version 4.3.1 was used for random forest analysis as a comparison to other AI models.

Results

We built numerous models including a decision tree and random forest model in order to create a classifier. A random seed is used to ensure that results are reproducible. We set the seed at 100. The number of trees built otherwise known as ntrees is 1000 trees within the model that we used. We first found the minimum lambda value for entry into the random forest model.

Lambda optimization

As shown in Fig. 1, lambda is the tuning parameter that can be used to decide how much we want to penalize the flexibility of our model. As the value of λ rises, there is a reduction in the value of coefficients resulting in the reduction of the variance, consequently avoiding overfitting. The value of lambda that gives the minimum mean cross-validated error—a vector of length(λ) = 0.034.

Random Forest classification

In Fig. 2, as an example, we ran this model using Random Forest (RF). The Random Forest model demonstrates that in the first split, 65% were less than 53 years of age. Out of the previous group, 31% is female and 34% is not female (male). Then, for a vast majority the model equated to 43 years of age. The numbers in each of the leaf nodes represent the samples for each of the nodes. For instance, the root node has 120 samples.

As seen in Fig. 3, the mean decrease in Gini impurity can be useful in many situations. For instance, the level of importance of each covariate can be determined by this impurity measure we calculated. According to the random forest regression findings—employment, depression, sleep trouble, gender, and diabetes status.

The random forest was determined to contain lambda that also gives the minimum mean cross-validated error—a vector of length(lambda) is 0.034. The Random Forest model performed better than all of the other models in the training dataset. According to Table 1, the random forest model had an accuracy of 98.4%. The other models included the null model which was comprised of a null binary logistic regression and had an accuracy of 87.8%. The decision tree classifier was 89.3% accurate. The generalized linear model was 87.8% accurate. Finally, the naïve Bayes model was 84.6%. This model is a supervised machine learning approach that classifies tasks. In comparison the null model (a model predicting the frequent class for the sample observation, classification accuracy of the random forest model was 98.4%.

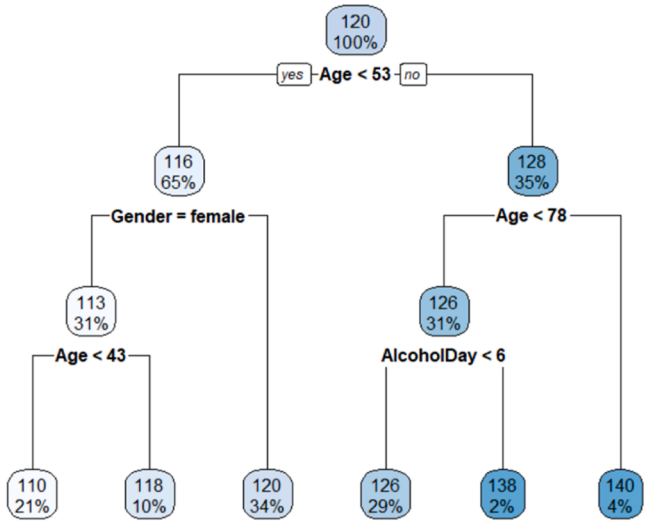

Fig. 2. Decision tree out of the Random Forest model explaining hypertension (root, split, and leaf nodes).

The results suggest that the random forest model that is being used only predicts the samples with 9% accuracy or number of wrongly classified observations, determined by the Out-of-Bag (a bootstrap aggregation method) error estimate. While making the samples, data points were chosen randomly and with replacement, and the data points which fail to be a part of that sample are known as out-of-bag points. In Table 2, we calculated that the confusion matrix yielded 86% accuracy.

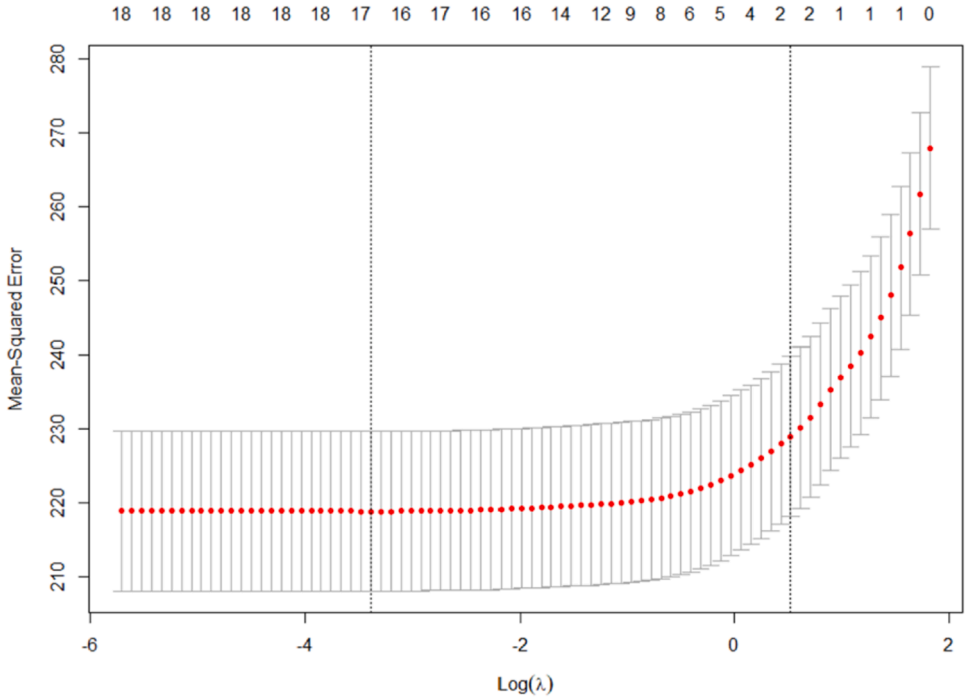

Fig. 1. Lambda optimization for a Random Forest Model with 95% confidence intervals.

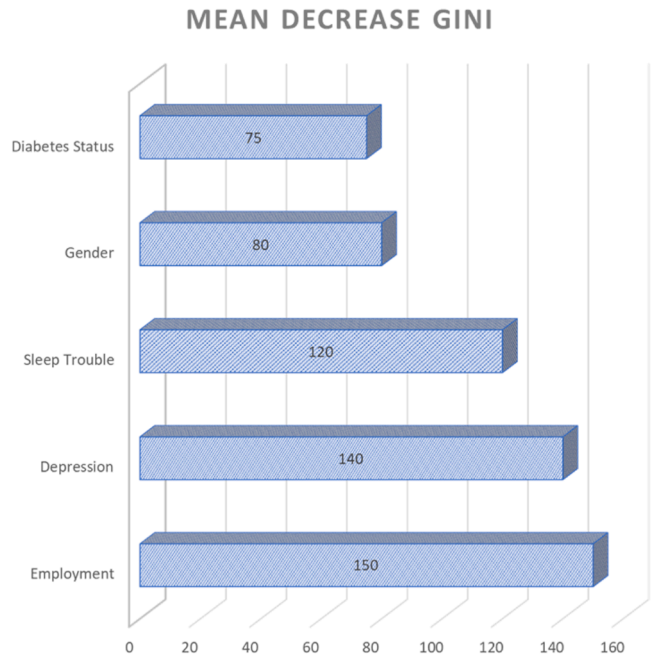

Fig. 3. Gini Coefficient shows the relative importance of each of the variables.

Table 1
Accuracy of each of the machine learning models.

Model	Train accuracy	Test Accuracy
Random Forest Model	98.4%	82.6%
Null Model (Logistic Regression)	87.8%	82.6%
Decision Tree Model	89.3%	82.6%
Generalized Linear Model	87.8%	82.6%
Naïve Bayes Model	84.6%	75.8%
Out-of-Bag Error	*9%*	

Table 2
Confusion matrix of the Random Forest Model.

	No	Yes	Class Error
No	1833	30	.02
Yes	154	105	.59

Discussion

In this study, by using real-life NHANES data, we found that among traditional AI models such as random forest modeling operate superior to other traditional machine learning techniques. Jackins et al. [10] found that for predicting health conditions such as diabetes, heart disease, and cancer, random forests were more accurate than Naïve Bayes. However, they did not compare the other machine learning models with Random Forest. Also, the findings in our study had used population data from the United States. The novel findings are that we were able to clearly demonstrate a relationship between physical health and mental health through using a unique AI model, known as Random Forest. We also found that using the random forest regressor model, the relative importance of employment was the most, as indicated by the Gini Coefficient. The next relatively important variables were as follows: depression, sleep trouble, gender, and diabetes status (in that order). These findings demonstrate that depression, after employment status, is the most important factor in predicting hypertension. Despite these practical applications of traditional machine learning techniques, LLM is expanding in leaps and bounds [9,10]. The challenge will be to understand how machine learning experts can creatively combine traditional and generative AI to create a synthesis and improved model for the future, being that this is the most challenging task in medical informatics. In this paper, we have revisited this question in the context of existing machine learning models. Due to the complex composition, a generalized method for classification will be needed.

Random Forest is a popular ensemble learning method that can handle classification tasks effectively. Previous researchers also have found Random Forest to be a relatively accurate Machine Learning model. While LLM uses neural networks and deep learning, random forests use machine learning. One way to combine LLM with Random Forest is to train a model using LLM sentence embeddings. These sentence embeddings are the numerical representations of sentences. Similarly, Yang et al. [11] found that using the two phases, pre-training and fine tuning, can be used for medical records. This technique can be used for collecting and abstracting data from medical charts. Then using state-of-the-art sentences, text, and image embeddings dataset, a random forest classifier can be created on the embedding vectors. The idea is to build combinations of multiple models created with various methods to combine strengths of each of the models and minimize weaknesses.

The Generative Pre-Trained Transformer (GPT) models, a type of LLM model, are ever-changing and are being trained on a massive amount of text data that can be measured through a growing number of parameters—sometimes in the billions [11–13]. For instance, GPT-3 has nearly 200 billion parameters. While this is 100 times larger than GPT-2, and two times the neurons in the human brain, this allows for the generation of text that varies contextually.

However, the models sometimes have unintended behaviors such as not following user instruction, creating facts, and generating biased text. One of the major uses is text prediction or predicting what are the next words in a text or email. Some of the limitations include the potential for this technology to perpetuate biases and stereotypes that comprise the original data. Stanford study indicates AI chatbots used by health providers are perpetuating racism. For example, misconceptions and falsehoods about Black patients, sometimes included fabricated, race-based equations [14,15]. Finally, when asked medical questions about kidney function, lung capacity, and skin thickness, the AI chatbot did not do well, posing the question of credibility. By having LLM recognize bias can create a more inclusive system.

Limitations of LLM

While Large Language Modeling is advancing at a rapid pace and has proven to be superior to existing AI models, numerous ethical questions arise. For instance, some misinformation can lead to rapid spread of more misinformation, perpetuating an infodemic. During the pandemic, we witnessed the rapid spread of misinformation leading to poor decision-making on the part of policymakers and individuals alike. Ethical considerations are so important that the European Union has created the European Ethical guidelines for trustworthy AI [16,17]. LLM can be considered a double-edged sword in that while LLM has opened the door to endless possibilities, one must proceed with caution. Some researchers have shown how rare conditions or disputed information often yields incorrect responses. Although known to have no connections, when asked if there is a relationship between SCN9A variants and epilepsy that is autosomal dominant. However, the LLM erroneously answered that there was a connection [6,18]. Biases emerging out of training models should be addressed adequately by creating machine learning models that can recognize biases. We must make sure that LLM is not accelerating disparities and perpetuating errors [19,20]. By adding diverse populations in training models, biased decision-making can be avoided.

Implications

As established in this study, the rapid spread of LLM is not making

other AI techniques obsolete. There are multiple ways that traditional machine learning techniques (Random Forest) can complement LLM to enhance the capabilities in AI. The datasets can be trained through LLM, then this can be applied to the Random Forest technique. LLM can be used for medical record abstraction regarding social determinants of health and summarizing charts. During decision-making, patient centered outcome measures can involve patients as prompted by LLM to create a brave new world of precision medicine. However, there are concerns regarding the training dataset being biased. This could negatively affect healthcare-related decision making by only using a certain select population rather than everyone. LLM is here to stay; however, before embracing this technology more research is needed about the potential biases. For instance, by training the model untrue information, this perpetuates biased information. Potential racial biases are possible by not accurately administering race-based questions [21]. However, in order to mitigate bias, researchers were able to train a Bidirectional Encoder Representations from Transformers (BERT) model to detect racial bias [21,22]. By harnessing LLM technology, there will be more time for healthcare workers to focus on the patient rather than having to go through the medical record in its entirety.

Conclusion

In conclusion, Random Forest undergirds many of the existing machine learning models. Instead of these being completely replaced, systems will learn to apply the advantages of LLM. For instance, digitized information such as medical records can easily be combined with depression screening in order to make clinical decisions based on random forest algorithms. As the findings from this study establish, depression relates to physical health. LLM can also be used to have a thorough understanding about rare diseases and potential genetic, phenotypic, and epigenetic research. The Electronic Health Records can be used to abstract information, predict 30-day readmissions, reduce lengths of stay in the hospital, and understand in-patient mortality using less training data in LLM. LLM and deep learning can enhance radiology imaging such as critical findings on an MRI. The purpose of LLM is to not to replace clinical judgment, but to augment judgment [23–25]. After all, allowing each technology to do what it does best should be the goal of optimal AI usage and application.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.G. Meyer, R.J. Urbanowicz, P.C.N. Martin, et al., ChatGPT and large language models in academia: opportunities and challenges, s13040-023-00339-9, BioData Min. 16 (1) (2023) 20, https://doi.org/10.1186/s13040-023-00339-9.

[2] T. Liu, X. Xiao, A framework of AI-based approaches to improving eHealth literacy and combating infodemic, Front. Public Health 9 (2021), 755808.

[3] C.K. Lo, What is the impact of ChatGPT on education? A rapid review of the literature, Educ. Sci. 13 (4) (2023) 410.

[4] M. Sallam, ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare, MDPI, 2023, p. 887. Accessed October 29, 2023, ⟨https://www.mdpi.com/2227-9032/11/6/887⟩.

[5] I. Seth, G. Bulloch, W.M. Rozen, Applications of artificial intelligence and large language models to plastic surgery research, Aesthet. Surg. J. 43 (10) (2023) NP809–NP810.

[6] C.M. Boßelmann, C. Leu, D. Lal, A.I. Are, language models such as ChatGPT ready to improve the care of individuals with epilepsy? Epilepsia 64 (5) (2023) 1195–1199, https://doi.org/10.1111/epi.17570.

[7] S. Lim, R. Schmälzle, Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering, Front. Commun. 8 (2023) 1129082.

[8] P.P. Ray, ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet Things Cyber-Phys. Syst. (2023). Accessed October 29, ⟨https://www.sciencedirect.com/science/article/pii/S266734522300024X⟩. Published online 2023. Accessed October 29.

[9] S. Montagna, S. Ferretti, L.C. Klopfenstein, A. Florio, M.F. Pengo, Data decentralisation of LLM-based chatbot systems in chronic disease self-management, in: Proceedings of the ACM Conference on Information Technology for Social Good, ACM, 2023, pp. 205–212. DOI: 10.1145/3582515.3609536.

[10] V. Jackins, S. Vimal, M. Kaliappan, M.Y. Lee, AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes, J. Supercomput. 77 (2021) 5198–5219.

[11] X. Yang, A. Chen, N. PourNejatian, et al., A large language model for electronic health records, NPJ Digit. Med. 5 (1) (2022) 194.

[12] J. Cabrera, M.S. Loyola, I. Magaña, R. Rojas, Ethical dilemmas, mental health, artificial intelligence, and LLM-based chatbots, in: I. Rojas, O. Valenzuela, F. Rojas Ruiz, L.J. Herrera, F. Ortuño (Eds.), Bioinformatics and Biomedical Engineering, Lecture Notes in Computer Science, Vol. 13920, Springer Nature Switzerland, 2023, pp. 313–326, https://doi.org/10.1007/978-3-031-34960-7_22.

[13] R. Davis, M. Eppler, O. Ayo-Ajibola, et al., Evaluating the effectiveness of artificial intelligence–powered large language models application in disseminating appropriate and readable health information in urology, J. Urol. 210 (4) (2023) 688–694, https://doi.org/10.1097/JU.0000000000003615.

[14] M. Cascella, J. Montomoli, V. Bellini, E. Bignami, Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios, J. Med. Syst. 47 (1) (2023) 33, https://doi.org/10.1007/s10916-023-01925-4.

[15] A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, L. Gutierrez, T.F. Tan, D.S. W. Ting, Large language models in medicine, Nat. Med. 29 (8) (2023) 1930–1940.

[16] J.A. Batsis, T.A. Mackenzie, R.T. Emeny, F. Lopez-Jimenez, S.J. Bartels, Low lean mass with and without obesity, and mortality: results from the 1999–2004 National Health and Nutrition Examination Survey, J. Gerontol. Ser. A Biomed. Sci. Med. Sci. 72 (10) (2017) 1445–1451.

[17] T. Lai, Y. Shi, Z. Du, et al.. Psy-LLM: Scaling up Global Mental Health Psychological Services with AI-based Large Language Models, 2023. Accessed October 29, 2023, ⟨http://arxiv.org/abs/2307.11991⟩.

[18] S.S. Biswas, Role of ChatGPT in public health, Ann. Biomed. Eng. 51 (5) (2023) 868–869, https://doi.org/10.1007/s10439-023-03172-7.

[19] Y. Liu, T. Han, S. Ma, et al., Summary of ChatGPT-related research and perspective towards the future of large language models, Meta-Radiology (2023), 100017.

[20] L.O. Gostin, J.G. Hodge, N. Valentine, H. Nygren-Krug, W.H. Organization, The domains of health responsiveness: a human rights analysis, Health Hum. Rights Work. Pap. Ser. (2023). Accessed October 29, ⟨https://apps.who.int/iris/bitstream/handle/10665/73926/HHRWPS2_eng.pdf⟩. Published online 2003. Accessed October 29.

[21] Are AI Language Models Such as ChatGPT Ready to Improve the Care of Individuals with Epilepsy? - Boßelmann - 2023 - Epilepsia - Wiley Online Library. ⟨https://onlinelibrary.wiley.com/doi/full/10.1111/epi.17570⟩ (Accessed October 30, 2023).

[22] European Commission and Directorate-General for Communications Networks, Content and Technology. Ethics Guidelines for Trustworthy AI, Publications Office, 2019, https://doi.org/10.2759/346720.

[23] S.Banerjee, G.M.Szirony, N.McCune, W.S.Davis, S.Subocz, B.Ragsdale. Transforming social determinants to educational outcomes: geospatial considerations. In: *Healthcare*. Vol 10. MDPI; 2022:1974. Accessed November 6, 2023. https://www.mdpi.com/2227-9032/10/10/1974.

[24] B Meskó, EJ Topol, The imperative for regulatory oversight of large language models (or generative AI) in healthcare, npj Digital Medicine 6 (1) (2023) 120.

[25] Liu Z, Zhong A, Li Y, et al. Radiology-GPT: A Large Language Model for Radiology. Published online June 14, 2023. Accessed November 6, 2023. http://arxiv.org/abs/2306.08666.