Improving Neurology Clinical Care With Natural Language Processing Tools

Wendong Ge, PhD,* Hunter J. Rice, BS, CRC,* Irfan S. Sheikh, MD,* M. Brandon Westover, MD, PhD, Allison L. Weathers, MD, Lyell K. Jones, MD, and Lidia Moura, MD, PhD, MPH

Correspondence
Dr. Moura
lidia.moura@
mgh.harvard.edu

Abstract

The integration of natural language processing (NLP) tools into neurology workflows has the potential to significantly enhance clinical care. However, it is important to address the limitations and risks associated with integrating this new technology. Recent advances in transformer-based NLP algorithms (e.g., GPT, BERT) could augment neurology clinical care by summarizing patient health information, suggesting care options, and assisting research involving large datasets. However, these NLP platforms have potential risks including fabricated facts and data security and substantial barriers for implementation. Although these risks and barriers need to be considered, the benefits for providers, patients, and communities are substantial. With these systems achieving greater functionality and the pace of medical need increasing, integrating these tools into clinical care may prove not only beneficial but necessary. Further investigation is needed to design implementation strategies, mitigate risks, and overcome barriers.

Introduction

Evidence-based medicine places data at the heart of decision-making processes. In modern practice, neurologists take in details of the patient's history, findings from examinations (e.g., muscle weakness, cognitive impairment), and results of ancillary tests (e.g., EEG, MRI). They then use all these data points in combination with their training, medical literature, and experience to formulate a diagnosis and care plan. These processes mandate a uniquely human connection between the physician and patient. Despite this connection, there are multiple places where the overall clinical workflow can be improved by technology, especially in the setting of data analysis and processing.

Recent advances in artificial intelligence (AI) such as those seen in Open-AI's ChatGPT have generated tremendous interest for enhancing industries from marketing to education.[1] Similarly, health care and more specifically, neurologic care should investigate the benefits and risks this and similar systems might provide. These tools might not only transform patient data for the benefit of both provider and patient, but they may present avenues for removing barriers that have built up in modern practice. In addition, these tools can allow for rapid and continual learning from large datasets, which makes them ideal for identifying complex symptoms in charts and investigating treatment efficacy.

With increasing levels of physician burnout, constrained access to neurologic care, and growing demand for neurology services, it is vital to explore ways of improving provider and patient experiences with neurology clinical care.[2,3] We discuss transformer-based natural language processing (NLP) models including understanding of their technology, identifying potential uses in neurology and neurosciences, highlighting challenges and limitations, recommendations

---

Glossary

of AI utilization and implementation, risks and barriers of AI, and the testing and validation needed for implementation into neurology health care models.

Recent Advancements: Technology and Terminology

There are various tiers by which technology in health care is taxonomized. Technology is "assistive" when it can detect clinically relevant data without analysis and generally requires the physician to finalize interpretation. "Augmentative" technology is when the machine analyzes or quantifies data in a clinically meaningful way but still requires interpretation by a physician or other qualified health care professional. And last, any machine that interprets data and independently generates clinically meaningful conclusions without physician involvement can be considered as "autonomous."[4]

To understand NLP and their functionality, it is important to understand their basic technology and terminology including machine learning, neural networks, and transformers (see eAppendix 1, links.lww.com/WNL/D179).

Limitations and Barriers to NLP in Neurologic Care

Before discussing the potential benefits these technologies pose for neurologic care, it is important to understand the challenges, limitations, and barriers that exist across various use cases. Understanding these limitations is crucial to ensure the safe and effective deployment of these technologies.

Patient Safety

Augmenting neurology clinical practice comes with risks to patient safety, confidentiality, and equity. NLP models are essentially large language models (LLMs), which means they use context to predict the next word and then generate whole sentences or paragraphs.[5] One weakness of LLMs is that they can combine plausible words to yield confabulation.

For neurology care, this inaccuracy can affect crucial medical information and treatment plans. In Figure 1A, we show an example whereby ChatGPT recommends valproic acid for a 25-year-old woman with juvenile myoclonic epilepsy (JME), who is planning for pregnancy. Although valproic acid is typically the drug of choice for patients with JME, it may not be considered for a 25-year-old woman in her reproductive years. Clinical experience suggests that lamotrigine, which is

also used in JME and has considerably less side effects than valproic acid, especially in pregnant women, would be better suited. This response generated did not consider the potential teratogenic effects of valproic acid in this age group.

Furthermore, NLP tools are limited in how clinical suggestions relate to the patient. As shown in Figure 1B, ChatGPT will often return vague or generalized responses when it does not possess enough information to make a definitive suggestion. In this case, various treatments have been cited in the literature as alternatives to epilepsy surgery, and yet options such as psychotherapies may be impractical for certain patients.

Until this limitation across multiple NLP models, including ChatGPT, can be resolved and rigorously validated, utilization of "over-the-counter" open-AI language model advises a cautionary approach.

Data Quality, Variability, and Security

Another limitation is quality and variability of the data. Inaccurate or incomplete data can affect the performance of machine learning models and algorithms. Data sources may vary in terms of format, structure, and quality, posing challenges for standardization efforts, documentation efficiency tools, and analysis techniques. Ensuring data accuracy, consistency, and representativeness remains a challenge in all use cases.

Incorporating this technology into clinical practice brings with it an innate risk for data security. To incorporate models into health care, these systems will need to be rigorously tested before any patient data is used for training and testing. There are risks surrounding privately owned closed-source nonmedical LLMs (ChatGPT, BERT, etc) because information ingested by these LLMs could violate data privacy laws or result in imprudent or unauthorized disclosure of confidential information. The standard licensing terms of these private platforms could include granting intellectual property rights in data submitted to the vendor making available the Generative AI, which would conflict with internal health system policies. It is therefore imperative that algorithms operate in end-to-end encryption and Health Insurance Portability and Accountability Act (HIPAA)–compliant systems.

Discrimination and Bias

Unlike humans, AI language models do not have personal opinions, beliefs, or biases. However, it is possible for models to generate biases that are present in the training data. This

Figure 1 Limitations of ChatGPT in Neurology Clinical Practice

A – Potentially Dangerous Advice

Prompt for ChatGPT:

What is the first line drug to use in a 25-year-old female patient who is in her reproductive years and is planning pregnancy and has just been diagnosed with juvenile myoclonic epilepsy (JME)?

ChatGPT Response:

The first-line drug of choice for a 25-year-old female patient with juvenile myoclonic epilepsy (JME) who is in her reproductive years and planning pregnancy is typically valproic acid (also known as sodium valproate).

Valproic acid has been shown to be effective in controlling seizures in the majority of individuals with JME and has a relatively low risk of teratogenicity (causing birth defects) compared to other antiepileptic drugs (AEDs). However, valproic acid is still associated with increased risk of neural tube defects and other major congenital malformations, and it is therefore recommended that women of childbearing age use effective contraception while taking valproic acid and discuss their individual risk with their doctor.

In general, the management of epilepsy in women of childbearing age requires a careful consideration of the risk-benefit ratio of the different AESs, as well as an assessment of the person's individual medical and seizure history. The choice of medication, as well as the starting does and titration schedule, will depend on several factors, including the type and frequency of seizures, the person's age and overall health, and any other medical conditions they may have.

It is important for women with epilepsy who are planning pregnancy to consult with their doctor and a specialist in epilepsy care, such as a neurologist or an epilepsy specialist, for guidance on the best treatment plan for their individual needs. This can help ensure that their seizures are well-controlled and that any potential risks to the developing fetus are minimized.

B – Vague/Inappropriate Advice

Prompt for ChatGPT:

How to surgically treat a patient with focal epilepsy with onset in the right mesial temporal region if they do not want surgery?

ChatGPT Response:

If a patient with focal epilepsy with onset in the right mesial temporal region does not want to undergo surgical treatment, alternative options should be considered. These options may include:

1. Continued medical management: The patient can continue to be treated with antiepileptic medications, which may provide adequate seizure control.
2. Non-invasive therapies: Alternative treatments such as neurostimulation techniques, including vagus nerve stimulation or transcranial magnetic stimulation, may be considered for patients who are not candidates for surgery or who do not want surgical treatment.
3. Psychotherapy: The patient may also benefit from psychotherapy to help them manage their epilepsy and related symptoms, such as anxiety or depression.

Ultimately, the best course of action will depend on the individual patient and their specific needs and circumstances. The decision to undergo surgery or consider alternative treatments should made in collaboration with the patient's doctor and based on a thorough evaluation of the patient's medical history and current health status.

Prompt conveys standard neurology clinical inquiry. ChatGPT Response is generated by ChatGPT.[18]

usually stems from biased language, stereotypes, or discriminatory attitudes that humans consciously or unconsciously imposed on the training dataset, which is then reflected in the model's output.[6] It is therefore imperative that models use diverse datasets and team's employee programmers from different perspectives, cultures, sexes, races, and ethnicities. It is the responsibility of developers, researchers, and organizations to be mindful of these issues and regularly evaluate them for biases. Future research would be needed prioritizing ethical considerations throughout the design, development, and deployment of advanced technologies. The design of this research should be focused on ensuring patient privacy, data security, and compliance with regulations through the implementation of robust privacy protocols, anonymization techniques, and secure data storage.

In addition, AI systems can only see outcomes as probabilities and trends. The interpretation of these findings is left to health care professionals. A major risk with incorporating these tools into health care is how it might lead to over medicalization or increased health disparities. It will come down to medical teams and governing bodies to scrutinize any AI medical findings before making guidance decisions. Furthermore, it is important that care teams and governing bodies remain well versed in how these models operate and their innate limitations. They must remain cognizant of the narrow utility of these tools and ensure they do not expand beyond their designated purpose or replace medical sense.

Training the "AI-NLP Team Member"

It is worth mentioning the effect of having an "AI-NLP Team Member." There would essentially need to be a neurology

residency for AI-NLP technology, and the efforts to train the assistant need to be balanced against the benefits of having such assistance. For very specialized task, there must be continued "training or retraining" of the models that would be required including updating them, revalidating and retesting as performed with any other software tool. Similar to the validation of other software's, there will need to be a specialized field of knowledge for the model. Nonetheless, continuing medical education for AI-NLP tools will have to be formalized and validated. The concept is similar to various proposals to add human Advanced Practice Providers to clinical workflow. Qualitative studies assessing the feasibility of training these "AI-NLP Team Members" will be needed to successfully implement them in neurology health care practices.

Barriers for Implementation

Several barriers are to be considered even before implementation of NLP models in Neurology clinics and departments. From a technological standpoint, NLP tools will need to overcome cost, time, and skill set needed to build, test, validate, and update current AI algorithms. For example, ChatGPT is a massive model with more than 175 billion parameters, which requires an enormous corpus for pretraining.[5,7] Because the information is stored within the whole model, it would be expensive and time-consuming to frequently update it. Unlike webpages or databases where information can be quickly changed, these complex models need to be fully retrained and fine-tuned. Some companies have started charging for premium access (higher limits, faster performance) to meet these financial constraints and finance future updates.[8] In addition, the timescale for retraining these models might work for well-established neurology guidelines and medical facts, but they would quickly become outdated with new publications and treatment innovations. Last, even dialog-capable systems, such as ChatGPT, will require machine learning experts and prompt engineers to maintain and update clinical tools.

Additional Limitations and Barriers

Additional limitations and barriers to NLP models in neurologic practice include factors of generalizability and external validation, cost and resource implications, the human-machine collaboration, and factors of interpretability and explainability (eAppendix 1, links.lww.com/WNL/D179).

Recommendations for AI-NLP Implementation and Utilization in Neurologic Care

Several recommendations can be made to optimize the implementation and utilization of advanced technologies in neurology clinical practice. By implementing these recommendations, a neurology clinical practice can leverage the potential of advanced NLP tools while addressing the limitations and barriers discussed earlier. These steps will contribute to safer, more effective, and ethically responsible deployment of emerging and new solutions, ultimately improving patient outcomes and enhancing the overall quality of neurology care.

Training and Education

To implement and use AI-NLP models into clinical practice, it is important to provide a comprehensive training and education program for clinicians to effectively use and understand the functionalities and limitations of these models. This includes front-end speech and ambient clinical intelligence (ACI). Offering continuous professional development opportunities is important to keep clinicians updated with the latest advancements and ensure they can leverage these technologies optimally.

Testing, Validation, and External Review

Testing and validation plays a crucial role in the development and deployment of language and machine learning (LLM) models, ensuring their accuracy, reliability, and suitability for real-world applications.

One key component to the validation of LLM is dataset selection, where careful consideration is given to choosing diverse and representative datasets that accurately reflect the target population and encompass various linguistic nuances and challenges.[9]

Benchmarking against existing models is another critical component of testing and validation. Comparing the performance of LLMs against state-of-the-art existing models is a way to achieve improvements and provide added value. This step helps establish the effectiveness and superiority of models, boosting their credibility and reliability in practical applications. Similar benchmarks have been investigated in medical imaging datasets.[10]

In addition, the establishment of robust evaluation metrics is imperative. These metrics serve as objective measures to assess the performance of LLMs, allowing for quantitative comparisons and meaningful interpretations. For implementation into clinical practice, it is prudent to carefully define and select appropriate evaluation metrics and take into account factors such as precision, recall, accuracy, and F1 score, among others, to ensure a comprehensive framework.

Moreover, the involvement of expert clinicians in the validation process is crucial. Their domain expertise and clinical insights are invaluable for evaluating the practical utility and effectiveness of LLMs in real-world health care settings. Collaborating with expert clinicians allows for critical feedback, validation of the clinical relevance and accuracy of the models, and identification for areas of improvement.

Incorporating external validation is a vital step in testing and validation process. By collaborating with external partners and

Figure 2 "History Bot" Summarizes Chart for Provider

Prompt for ChatGPT:
Please briefly summarize this patient encounter and list highlights and areas of concern:

Entry: (Filed by provider)
64 right-handed male presents to the epilepsy clinic for first lifetime GTC activity.
Seizure onset: Patient went to sleep and at 3 AM wife reports witnessing his seizure.
Seizure semiology: Woke up from sleep with ictal vocalization followed by generalized tonic clonic activity (bilateral upper extremity arm clenching) and post-ictally was breathing heavily and drooling at the mouth. Seizure lasted for x3 minutes, post-ictally he was fatigued and confused, nonresponsive and started to breath heavily through his mouth. Taken by ambulance to local emergency department.
CTH: normal
MRI brain: 9 mm pituitary macroadenoma descended in the right side of the sella with some invasion into the right cavernous sinus without impingement or vascular structures. No acute infract, intracranial hemorrhage or mass lesion
Seizure risk factors: No family history of seizures, TBI/cracked skull at 7 years of age, fell down stairs, no seizures as a child, no children with seizures
Current ASM: Keppra 500/500
Patient states that he has been experiencing sensation of feelings of doom and feelings of doom that last for 30 sec and then disappear. Frequency would be every 2 months but hasn't experienced in last 6 months. This was thought to be situational as his father had just passed away from cancer. No issues with sleep, light snoring but no RBD. No issues with memory or cognition.
Past Medical/Surgical History: Arthritis of hip, Gastroesophageal reflux disease, Hyperlipidemia, Prostate cancer.
Allergies: No Known Allergies.
Family History: Colon polyps in his father; Prostate cancer in his father and paternal grandfather.
Social History: Does not smoke, drinks EtOH 2 glasses/night, no other recreational drugs at times. Retired but worked as a teacher. Married with 2 kids who are all healthy.
Review of Systems: General review of systems was negative for fevers, chills, rashes, change in weight, energy level or appetite, chest pain, palpitations, shortness of breath, cough, abdominal pain, nausea, vomiting, and change in bowel or bladder habits (i.e., incontinence).
Physical examination was non focal and vital signs were within normal range. Prior to admission, he was not taking any medications, either prescribed or over the counter

ChatGPT Response:
Highlights:
- 64-year-old male with first lifetime generalized tonic-clonic (GTC) seizure activity.
- Seizure onset while asleep with ictal vocalization and bilateral upper extremity arm clenching.
- Post-ictally was fatigued, confused, and breathing heavily.
- MRI brain revealed a 9 mm pituitary macroadenoma without impingement or vascular structures.
- Patient reports feeling of doom, but thought to be situational due to father passing away from cancer.
- No other significant medical history, surgical history, or family history of seizures.

Areas of concern:
- The presence of a pituitary macroadenoma may indicate an underlying cause for the patient's seizure activity and warrants further investigation.
- The patient's history of feeling of doom may be a potential prodromal symptom for seizures and should be monitored closely.
- It is important to evaluate the patient for any possible triggers for seizures and adjust his medication regimen accordingly.

Prompt conveys standard neurology clinical inquiry. Entry created by neurologist to emulate a real patient note. ChatGPT Response is generated by ChatGPT.[18]

leveraging independent datasets, one can assess the performance, generalizability, and real-world applicability of LLMs.[11] Rigorous testing and validation protocols help ensure that the models are fit for their intended purpose and mitigate the risks associated with their usage beyond their intended scope.

Ethical and Legal Considerations (Focus on NLP)

Implementing AI-NLP models requires careful consideration of ethical and legal factors. Ensuring patient privacy, data security, and compliance with regulations such as HIPAA and General Data Protection Regulation is essential. In addition, addressing potential biases in algorithms, maintaining transparency in data usage, and obtaining informed consent from patients are critical ethical considerations. Obtaining US Food and Drug Administration (FDA) clearance of AI-NLP medical devices is also essential before mass production and utilization into health care systems. No AI-NLP model has been approved by the FDA for usage in clinical practice to date.[12]

Recommendations for implementation include the prioritization of ethical considerations throughout the design, development, and deployment of advanced technologies. Implementing robust privacy protocols, anonymization techniques, and secure data storage can ensure patient privacy, data security, and compliance with regulations. Steps to mitigate biases and ensure fairness in algorithms too are essential to prevent disparities in health care delivery. Previous models have been offered to show the incorporation of AI into the legal standard of care for a clinical situation.[13]

Additional Recommendations for Implementation

Additional recommendations for NLP implementation within neurology clinical practice include collaboration with all

Figure 3 "AI Team Member" Gives Assessment and Care Options

Prompt for ChatGPT:
For the below patient, please provide an assessment, create a care plan, and provide an explanation for the assessment and plan:

Entry:
     *Entry from Figure 2

ChatGPT Response:
The patient is a 64-year-old male who presented to the epilepsy clinic for his first lifetime generalized tonic-clonic (GTC) activity. The seizure onset occurred during sleep and lasted for approximately 3 minutes. The patient exhibited ictal vocalization followed by bilateral upper extremity arm clenching and post-ictally was breathing heavily and drooling at the mouth. He was taken to the emergency department and was found to have a 9 mm pituitary macroadenoma that invaded the right cavernous sinus without impingement on vascular structures. His past medical history includes arthritis of the hip, gastroesophageal reflux disease, hyperlipidemia, and prostate cancer. He is currently taking Keppra 500/500 for his seizure. The patient has been experiencing feelings of doom that last for 30 seconds every 2 months, which is thought to be situational. His family history includes colon polyps in his father and prostate cancer in his father and paternal grandfather.

Care Plan:
1. Consult with a neurosurgeon to evaluate the pituitary macroadenoma and determine the best course of treatment.
2. Increase the dose of Keppra to 1000/1000 and schedule a follow-up appointment to monitor the patient's response to the medication.
3. Obtain an electroencephalogram (EEG) to evaluate the patient's brain wave activity and determine if there is any abnormal activity that could be contributing to his seizure.
4. Educate the patient and his family on seizure safety and precautions to take in case of future seizure activity.
5. Schedule a follow-up appointment in 6 months to evaluate the patient's seizure activity and adjust the treatment plan as necessary.

Explanation:
The patient's presentation is concerning for a new-onset seizure, which requires further evaluation and treatment. The presence of a pituitary macroadenoma may be contributing to his seizure activity and requires consultation with a neurosurgeon for further evaluation and treatment. Increasing the dose of Keppra may help control the patient's seizure activity, but follow-up is necessary to monitor his response to the medication. Obtaining an EEG can help determine if there is any abnormal activity in the brain that may be contributing to his seizure activity. Educating the patient and his family on seizure safety and precautions is important to ensure their safety in case of future seizure activity. A follow-up appointment in 6 months will allow for evaluation of the patient's seizure activity and adjustment of the treatment plan as necessary.

Prompt conveys standard neurology clinical inquiry. Entry is same as seen in Figure 2 entry. ChatGPT Response is generated by ChatGPT.[18]

stakeholders, having standardization and interoperability and considering ways for continuous improvement and iterative development for these models (eAppendix 1, links.lww.com/WNL/D179).

Applications of NLP in Neurology Clinical Practice

The integration of NLP tools in neurology clinical practice holds significant promise for enhancing patient care and advancing research.[14,15] These tools have the potential to streamline the analysis of clinical notes, other electronic health records, and medical literature, allowing neurologists to efficiently extract relevant information and gain deeper insights into patient conditions.[16] NLP can aid in automating the extraction of patient demographics, medical history, symptoms, and diagnostic information from textual data, enabling improved clinical decision-making and personalized treatment approaches.[17] Furthermore, NLP tools can assist in the identification and monitoring of disease progression, detecting patterns in large-scale datasets, predicting treatment outcomes, and identifying adverse effects.[18] In addition, NLP can support the creation of neurology-specific ontologies and standardized vocabularies, facilitating better information exchange and interoperability across different health care systems and research institutions.[19] Applications of NLP models can also be expanded to inpatient and outpatient operations and be suitable in the creation of new data streams (eAppendix 1, links.lww.com/WNL/D179). The below discussion entails incorporations of NLP tools in clinical practice, which if entertained in future neurology health care systems would most certainly require validation through rigorous prospective and randomized studies.

Chart Review and Documentation

To demonstrate how NLP models could potentially affect neurologic care, we provide the following simulated case as an example. Consider a 64-year-old man with new-onset generalized tonic-clonic seizures, who presented a week ago to an

emergency department (ED). The patient is now being seen in an outpatient neurology clinic after receiving MRI and CT imaging. During the appointment, the provider will need to evaluate risk factors, address the recent ED visit, review imaging results, assess neurologic status, and create an assessment and personalized care plan.

Before the neurologist even starts this new patient consultation, they will need to familiarize themselves with information taken from the ED visit and the imaging results. Typically, this will include a detailed chart review, but in a busy clinic, an NLP tool might be able to augment this process while ensuring the provider reviews important items and concerns. The "History Bot" in Figure 2 shows how the notes for a hypothetical patient might be summarized using an AI-based NLP model. In this case, we had used open-AI.

Once the neurologist is ready, they would then start the consultation and document the patient's history of present illness including medical history, medications, and social history. This is another area where NLP can enhance the visit: advancements in medical documentation to improve clinician-patient experiences. Cutting-edge technologies such as front-end speech and AI-driven ACI have further accelerated the documentation process.[20] Clinicians can now dictate patient encounters directly into electronic medical records using voice commands, significantly improving efficiency. ACI captures patient encounters through mobile applications and automatically converts them into comprehensive medical records, enhancing accuracy and streamlining the workflow. These advancements, in conjunction with conversational and ambient AI, hold potential for further automation and simplification of clinical tasks, paving the way for a future where health care is driven by innovative technologies.

An example of this is the NLP-enabled speech-to-text systems such as Nuance's Dragon Ambient eXperience and 3M's M*Modal Fluency Direct; this consultation can be directly transcribed into the electronic health record.[20] In March and April 2023, partnerships were announced between Nuance and Microsoft[21] HCA Healthcare and Augmedix[22] and Amazon and 3M[23] for the development of such clinical documentation systems.

Of note, these speech-to-text systems are not without drawbacks. One apparent one is the prone to error. There is a need to balance the ease of generating the summary with the necessity to double-check it. Just as with human scribes, it may take longer to edit a file created by AI than to gather and document the information oneself. Further studies comparing independent documentation, human scribes, and NLP-enabled speech-to-text systems will be needed before AI can be considered to be completely autonomous.

Assessment and Care Plans

NLP tools may also help providers formulate thorough assessments and suggest care plans. Unlike current internet medical services, which search through restrictive symptom lists for potential diagnoses, NLP tools could potentially search through a

patient's entire chart to identify key aspects of their history and test results. Furthermore, by implementing these tools in neurology clinics, they could enhance a trained physician's ability to problem solve complex cases and formulate differential diagnoses in much the same way a medical team might collaborate. As shown in Figure 3, an "AI Team Member," using ChatGPT for demonstration purposes, could use NLP to identify key aspects of our patient's chart and provide customized recommendations. Furthermore, these care options come from the vast corpus of medical knowledge that the model is trained on. By increasing and updating the dataset used to train these models, these AI Team Members could improve their suggestions and provide alternative and cutting-edge treatment options.

Once a care plan is chosen, the "AI Team Member" could also help identify medication conflicts, potential side effects, or more cost-effective care options. BERT-based systems have recently been developed to identify adverse drug reactions using current data.[24] Furthermore, by giving these tools access to current drug prices and insurance information, they could provide patient-specific treatment costs (eAppendix 1, links.lww.com/WNL/D179). These tools could also be programmed with information specific to the health system and local community including community resources and support groups and financial assistance plans.

Medical Education

Patients and caregivers might also benefit from diagnostic and treatment-related information. By providing a well-trained and validated AI chat service to communities, health care systems can ensure patients get their questions answered quickly and accurately.[25] Easy-to-understand health information would give patients greater autonomy and confidence in their care. For now, some AI-NLP models, especially ChatGPT, include a disclaimer for clinical questions that reminds patients to speak with a licensed health care provider.[5] However, a reviewed and validated tool designed and tested by a hospital can direct patients to their own care team for future questions.[26] These tools can also attract new patients to a health system/practice.

In addition, per the 21st Century Cures Act, which restricts information blocking, patients now have unprecedented and immediate access to their medical records. While there are myriad benefits to this transparency, instances of misunderstandings, confusion, and anxiety are bound to continue, and AI-NLP tools could help with understanding and even simplifying medical terminology and connect patients and their families to verified resources.

Conclusion

As we advance further into the 21st century, we believe AI-NLP models have the potential to overlap with health care systems. Several questions are yet to be answered however: "Can AI-NLP be autonomous?" and "When will neurology health care systems be ready to consider AI-NLP as an

integral part of the team?" The answer to these questions would require prospective and randomized studies for validation and additional studies needed on how to train and assess the performance of these AI tools, similar to how medical students partake in the United States Medical Licensing Examination or residents/fellows must meet annual Accreditation Council for Graduate Medical Education milestones.

Using AI-NLP tools in clinical practice could help neurology clinics addresses the increasing demand for neurology care.[27] As this technology continues to improve, doubts about its efficacy will only erode further. However, the benefits of integrating this technology for patients, providers, and communities must be weighed against inherent risks. Governing bodies and vendors should begin to design testing strategies and safe points of integration for these tools. Meanwhile, neurologists and clinical staff should work with programming teams to design workflows and implementation strategies. But most importantly, patients and communities need to be placed front and center. Their data security, care, and trust should be all parties' top priority. By starting this process now, we not only equip neurologists with the tools they need to deliver the highest quality care but allow for the creation of safe and effective solutions.

Study Funding

Disclosure

Appendix Authors

Name	Location	Contribution
Wendong Ge, PhD	Department of Neurology, Massachusetts General Hospital, Boston	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Hunter J. Rice, BS, CRC	Department of Neurology, Massachusetts General Hospital, Boston	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Irfan S. Sheikh, MD	Department of Neurology, Massachusetts General Hospital, Boston	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
M. Brandon Westover, MD, PhD	Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA	Drafting/revision of the article for content, including medical writing for content; study concept or design; and analysis or interpretation of data
Allison L. Weathers, MD	Information Technology Division, Cleveland Clinic, OH	Drafting/revision of the article for content, including medical writing for content; study concept or design; and analysis or interpretation of data
Lyell K. Jones, MD	Department of Neurology, Mayo Clinic, Rochester, MN	Drafting/revision of the article for content, including medical writing for content; analysis or interpretation of data
Lidia Moura, MD, PhD, MPH	Department of Neurology, Massachusetts General Hospital; Department of Neurology, Harvard Medical School, Boston, MA	Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

References

1. Villasenor J. How ChatGPT Can Improve Education, Not Threaten It. Scientific American; 2023.
2. Weathers AL. Chipping away at neurologist burnout, one refill request at a time. *Neurol Clin Pract.* 2016;6(5):379-380. doi:10.1212/CPJ.0000000000000292
3. Yates SW. Physician stress and burnout. *Am J Med.* 2020;133(2):160-164. doi:10.1016/j.amjmed.2019.08.034
4. Frank RA, Jarrin R, Pritzker J, et al. Developing current procedural terminology codes that describe the work performed by machines. *NPJ Digit Med.* 2022;5(1):177. doi:10.1038/s41746-022-00723-5
5. OpenAI. GPT-4 Technical Report. 2023.
6. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med.* 2018;178(11):1544-1547. doi:10.1001/jamainternmed.2018.3763

7. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving Language Understanding by Generative Pre-Training. 2018. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

8. Wiggers K. *OpenAI Begins Piloting ChatGPT Professional, a Premium Version of Its Viral Chatbot*. TechCrunch; 2023.

9. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. *Patterns (NY)*. 2021;2(10):100347. doi: 10.1016/j.patter.2021.100347

10. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *N Engl J Med*. 2023;388(21):1981-1990. doi:10.1056/NEJMra2301725

11. Lin Y, Sharma B, Thompson HM, et al. External validation of a machine learning classifier to identify unhealthy alcohol use in hospitalized patients. *Addiction*. 2022; 117(4):925-933. doi:10.1111/add.15730

12. US FDA. Artificial Intelligence and Machine Learning (AI/ML): Enabled Medical Devices. 2022. Accessed July 15, 2023. fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

13. Price WN II, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA*. 2019;322(18):1765-1766. doi:10.1001/jama.2019.15064

14. Xie K, Gallagher RS, Conrad EC, et al. Extracting seizure frequency from epilepsy clinic notes: a machine reading approach to natural language processing. *J Am Med Inform Assoc*. 2022;29(5):873-881. doi:10.1093/jamia/ocac018

15. Tyagi T, Magdamo CG, Noori A, et al. NeuraHealth: an automated screening pipeline to detect undiagnosed cognitive impairment in electronic health records with deep learning and natural language processing. 2022. ArxIV. Preprint posted online June 20, 2022. https://arxiv.org/pdf/2202.00478.pdf.

16. Harkema H, Chapman WW, Saul M, Dellon ES, Schoen RE, Mehrotra A. Developing a natural language processing application for measuring the quality of colonoscopy procedures. *J Am Med Inform Asscoc*. 2011;18(suppl 1):i150-i156. doi:10.1136/amiajnl-2011-000431

17. Berge GT, Granmo OC, Tveit TO, Munkvold BE, Ruthjersen AL, Sharma J. Machine learning-driven clinical decision support system for concept-based searching: a field trial in a Norwegian hospital. *BMC Med Inform Decis Mak*. 2023;23(1):5. doi:10.1186/s12911-023-02101-x

18. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. *Patterns (NY)*. 2021;2(10):100347. doi: 10.1016/j.patter.2021.100347

19. Azizi S, Hier DB, Wunsch Ii DC. Enhanced neurologic concept recognition using a named entity recognition model based on transformers. *Front Digit Health*. 2022;4: 1065581. doi:10.3389/fdgth.2022.1065581

20. Ambient Clinical Intelligence | Automatically Document Care | Nuance. Nuance Communications. 2023.

21. Keating C. Nuance and Microsoft Announce the First Fully AI-Automated Clinical Documentation Application for Healthcare. Nuance Communications; 2023.

22. Chesler M. Augmedix Announces Partnership with HCA Healthcare to Accelerate the Development of AI-enabled Ambient Documentation. *GlobeNewswire Newsroom*. April 18, 2023.

23. Source 3M. 3M Health Information Systems collaborates with AWS to accelerate AI innovation in clinical documentation. *PR Newswire Cision*. April 18, 2023. Accessed October 28, 2023. https://www.prnewswire.com/news-releases/3m-health-information-systems-collaborates-with-aws-to-accelerate-ai-innovation-in-clinical-documentation-301799571.html.

24. Martenot V, Masdeu V, Cupe J, et al. LiSA: an assisted literature search pipeline for detecting serious adverse drug events with deep learning. *BMC Med Inform Decis Mak*. 2022;22(1):338. doi:10.1186/s12911-022-02085-0

25. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer*. 2021;7(4):e27850. doi:10.2196/27850

26. Pieces Predict. Accessed July 15, 2023. piecestech.com.

27. DeBusk A, Subramanian PS, Scannell Bryan M, Moster ML, Calvert PC, Frohman LP. Mismatch in supply and demand for neuro-ophthalmic care. *J Neuroophthalmol*. 2022; 42(1):62-67. doi:10.1097/WNO.0000000000001214