DISKUSSION/KOMMENTAR/PERSPEKTIVEN

Die Rolle von erklärbarer KI und Evaluierungsrahmen für die sichere und effektive Integration großer Sprachmodelle im Gesundheitswesen

Sandeep Reddy, MBBS, MSc, PhD[1] ; Alexandre Lebrun, MS, MSEE[2] ; Adam Chee, PhD[3] ; und Dimitrios Kalogeropoulos, PhD[4,5]

[1]Direktor, Deakin School of Medicine, Victoria, Australien; [2]Mitbegründer und CEO, Nabla, Paris, Frankreich; [3]Außerordentlicher Professor, Saw Swee Hock School of Public Health, National University of Singapore, Singapur; [4]Gründer und Chief Executive, Global Health & Digital Innovation Foundation, London, UK; [5]UCL Global Business School for Health, London, UK

Korrespondierender Autor: Sandeep Reddy; E-Mail: sandeep.reddy@deakin.edu.au DOI:

https://doi.org/10.30953/thmt.v9.485

Schlüsselwörter: KI, künstliche Intelligenz, Converge2Xcelerate, Gesundheitswesen, große Sprachmodelle, LLM

Eingereicht: 8. Februar 2024; Angenommen: April 2, 2024; Veröffentlicht: April 30, 2024

Die Integration von künstlicher Intelligenz (KI), insbesondere von großen Sprachmodellen (LLMs: eine spezielle Art von KI, die auf großen Textmengen trainiert wird)
um vorhandene Inhalte zu verstehen und originäre Inhalte zu generieren) in das Gesundheitswesen nimmt weiter zu und erfordert eine sorgfältige Bewertung und Überwachung, um einen sicheren, ethischen und effektiven Einsatz zu gewährleisten. Während einer Podiumsdiskussion auf der Converge2Xcelerate-Veranstaltung 2023, die vom Herausgeber dieser Zeitschrift, Partners in Digital Health, organisiert wurde, tauschten sich KI-Experten über die Chancen und Herausforderungen aus, die sich bei der durchdachten Umsetzung von Innovationen wie LLM im Gesundheitswesen ergeben.[1]Eine Zusammenfassung der wichtigsten Perspektiven aus der Podiumsdiskussion wird hier vorgestellt.

Die wichtigsten Themen sind:

- Das Potenzial erklärbarer KI zur Förderung von Transparenz und Vertrauen;
- Herausforderungen bei der Abstimmung von KI mit variablen globalen Gesundheitsprotokollen;
- Die Bedeutung der Evaluierung durch translationale und Governance-Rahmenwerke, die auf den Kontext des Gesundheitswesens zugeschnitten sind;
- Skepsis gegenüber einer zu weit gehenden Verwendung von LLMs für Dialogschnittstellen;
- die Notwendigkeit, LLMs unter Berücksichtigung des Risikoniveaus mit Bedacht zu validieren.

Darüber hinaus wurde in der Diskussion die Erklärbarkeit hervorgehoben (d. h. das Konzept, dass ein maschinelles Lernmodell und seine (d. h. das Konzept, dass ein Modell des maschinellen Lernens und seine Ergebnisse erklärt werden können und für einen Menschen auf einem akzeptablen Niveau "Sinn ergeben"), die Bewertung und die sorgfältige Abwägung mit den Fachleuten im Gesundheitswesen als entscheidend für die Realisierung von Vorteilen bei gleichzeitiger proaktiver Bewältigung der Risiken einer größeren Einführung von KI in der Medizin. Andere Diskussionsthemen konzentrierten sich auf kritische Bewertungspraktiken für KI in der Medizin, die Notwendigkeit und die Grenzen erklärbarer KI und die Überlegungen, die Führungskräfte im Gesundheitswesen vor dem Einsatz anstellen müssen.

Der Einsatz von KI-Techniken, insbesondere von LLMs wie GPT-4 (Generative Pre-trained Transformer 4), in der Verwaltung und in der klinischen Praxis des Gesundheitswesens beschleunigt sich rapide.[2]Die Integration von KI verspricht zwar, die menschlichen Fähigkeiten zu erweitern und den Zugang, die Qualität und die Effizienz zu verbessern, bringt aber auch komplexe technische, ethische und regulatorische Überlegungen in Bezug auf die Transparenz sowie die Rechenschaftspflicht und die Auswirkungen auf Patienten und Fachkräfte im Gesundheitswesen mit sich.[3,4]

Erklärbare KI: Vertrauen und Transparenz schaffen

Ein Schwerpunkt des Gesprächs war die Rolle von erklärbaren KI-Techniken bei der Schaffung von Vertrauen in KI-Systeme. Die Teilnehmer stimmten darin überein, dass LLMs zwar nützliche Ergebnisse liefern können, aber von Natur aus als "Black Boxes" fungieren, bei denen die Gründe für die Schlussfolgerungen nicht ersichtlich sind. Diese Undurchsichtigkeit wird im Gesundheitswesen, wo viel auf dem Spiel steht, zu einem Problem.

Die Diskussionsteilnehmer betonten, dass erklärbare KI ein aktives Forschungsgebiet ist, um diese Transparenzprobleme zu lösen.

Es erleichtert genaue und wiederholbare Ergebnisse und verdeutlicht gleichzeitig die Zusammenhänge zwischen Inputs und Outputs.[5]Die Realisierung einer umfassenden Erklärbarkeit mit großen neuronalen Netzen bleibt jedoch eine Herausforderung. Die komplexe, multivariate und dynamische Beschaffenheit klinischer Umgebungen kann Erklärungsansätze, die auf ein begrenzteres Umfeld abgestimmt sind, zusätzlich erschweren.[6]

Bewertung: Sicherstellung einer wirksamen und ethischen Übersetzung

Diese Beobachtungen unterstreichen die Notwendigkeit einer rigorosen, auf Anwendungen im Gesundheitswesen zugeschnittenen Evaluierung während der gesamten Entwicklung, Einführung und des Betriebs von KI-Systemen. Die Diskussionsteilnehmer sprachen sich für umfassende, kontinuierliche Evaluierungsrahmen wie TEHAI (Translational Evaluation of Healthcare AI)[7]aus, die die Integration in IT-Systeme des Gesundheitswesens, klinische Akzeptanzfaktoren, eine aktualisierte Leistungsüberwachung und Governance-Komponenten[8]wie Rechenschaftspflicht und Ethik umfassen. Die Diskussionsteilnehmer wiesen auch darauf hin, dass sich KI von den engeren regulatorischen Bewertungen von Sicherheit und Schaden allein unterscheidet. Eine multidimensionale Bewertung kann Stärken, Schwächen und geeignete Anwendungsfälle aufzeigen, um die Einführung von KI verantwortungsvoll zu steuern.

Automatisierung von Pflegeprozessen, nicht von Gesprächen

Eine weitere Überlegung, die während der Podiumsdiskussion aufgeworfen wurde, war die zunehmende Neigung zur Vermenschlichung von KI-Systemen, insbesondere von Konversationsschnittstellen. Die Diskussionsteilnehmer warnten davor, dass eine übermäßige Personifizierung zu unpraktischen Erwartungen hinsichtlich Vollständigkeit und Perfektion führt. LLMs können zwar dabei helfen, wichtige Pflegeprozesse zu automatisieren, aber eine vollautomatische Diagnose oder Behandlung ist derzeit noch unzuverlässig.

Darüber hinaus schlugen die Diskussionsteilnehmer vor, die Automatisierung auf die Unterstützung der Kliniker bei der Dokumentation, Kodierung und Pflegekoordination zu konzentrieren. Erklärungen zur künstlichen Intelligenz dienen eher dazu, geeignete mentale Modelle für ihre Fähigkeiten und Grenzen zu entwickeln, als sie zu vermenschlichen.[9]

Fortschritt durch Besonnenheit

Zusammenfassend betonten die Podiumsteilnehmer, dass vorsichtige, kollaborative[8]und evaluative Ansätze für eine produktive Anwendung von KI-Fortschritten in der klinischen Praxis unerlässlich sind. Anstatt sie zu ersetzen, liegt der Schwerpunkt darauf, mit unterstützenden, risikoärmeren Rollen zu beginnen und die Vorteile und Mängel, die durch den Einsatz in der realen Welt zutage treten, transparent zu bewerten.[7]

Kontinuierliche Fortschritte hängen von offenen und ehrlichen Bewertungen durch interdisziplinäre Führungskräfte ab, um die konstruktivsten Nischen für aufkommende Innovationen wie LLMs in der tatsächlichen Versorgung zu finden. Erklärbare Bewertungsrahmen, die die Transparenz erleichtern, können eine wesentliche Rolle dabei spielen, die Integration von KI verantwortungsvoll voranzutreiben.

Finanzierung

Für die Erstellung des Manuskripts wurden keine externen Mittel in Anspruch genommen.

Finanzielle und nicht-finanzielle Beziehungen und Aktivitäten

Sandeep Reddy ist Vorsitzender und Direktor von Healea Services und des Zentrums zur Förderung der translationalen KI in der Medizin.

Mitwirkende

Sandeep Reddy war für den Entwurf und die Endredaktion des Manuskripts verantwortlich. Alle Autoren haben zum Inhalt des Manuskripts beigetragen.

Datenverfügbarkeitserklärung (DAS), gemeinsame Nutzung von Daten, Reproduzierbarkeit und Datenrepositorien

Für die Erstellung dieses Artikels wurden keine Originaldaten verwendet.

Anwendung von KI-generiertem Text oder verwandter Technologie

KI-Technologie wurde eingesetzt, um das Transkript der Podiumsdiskussion zu erstellen und die Grammatik bei der Erstellung des Manuskripts zu unterstützen.

Danksagung

Die Autoren bedanken sich bei Partners in Digital Health für die Ausrichtung der Veranstaltung und der Podiumsdiskussion.

Referenzen

1. Telemedizin und Medizin heute. ConV2X-Symposium [Internet]. [zitiert 2024 Feb 1]. Verfügbar unter: https://telehealth.conv2xsym-posium.com
2. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Künstliche Intelligenz (KI) und globale Gesundheit: Wie kann KI zur Gesundheit in ressourcenarmen Gebieten beitragen? BMJ Glob. Health. 2018;3:e000798. https://doi.org/10.1136/bmjgh-2018-000798
3. Char DS, Shah NH, Magnus D. Implementing machine learn-ing in health care-addressing ethical challenges. N Engl J Med. 2018;378:981–3. https://doi.org/10.1056/NEJMp1714229
4. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Die wichtigsten Herausforderungen für die Erzielung klinischer Wirkung mit künstlicher Intelligenz. BMC Med. 2019;17. https://doi.org/10.1186/ s12916-019-1426-2
5. Adadi A, Berrada M. Peeking inside the black-box: a sur-vey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60. https://doi.org/10.1109/ACCESS.2018.2870052
6. Rudin C. Stop explaining black box machine learning mod-els for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206-15. https://doi.org/10.1038/ s42256-019-0048-x
7. Reddy S, Rogers W, Makinen VP, Coiera E, Brown P, Wenzel M, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ Health Care Inform. 2021;28(1):e100444. https://doi.org/10.1136/bmjhci-2021-100444

8. Kalogeropoulos D, Barach P. Telehealth's role enabling sustain-able innovation and circular economies in health. Telehealth Med Today. 2023;8(1). https://doi.org/10.30953/thmt.v8.409
9. Miller T. Erklärung in der künstlichen Intelligenz: Erkenntnisse aus den Sozialwissenschaften. Artif Intell. 2019;267:1-38. https://doi.org/10.1016/j.artint.2018.07.007