MedWiss.Online

Von Black Box zu Glasbox: Erklärbare KI in der Schlaganfallbehandlung

Das Forschungsprojekt „Liberate AI“ vereint Expertisen aus Medizin, Informatik und vertrauenswürdiger künstlicher Intelligenz, um ein KI-Modell zu entwickeln, das Ärzt:innen bei der Behandlung des ischämischen Schlaganfalls unterstützt. Als digitales Assistenzsystem soll es den langfristigen Behandlungserfolg einer mechanischen Thrombektomie sowie mögliche Komplikationen vorhersagen. Mithilfe des Swarm Learning wird das KI-Modell privatsphäreschonend mit medizinischen Daten trainiert, die an verschiedenen Standorten in Deutschland vorliegen. Liberate AI beschäftigt sich zudem mit der Erklärbarkeit der KI sowie ihrer Fähigkeit, differenzierte Vorhersagen für Patientensubgruppen zu treffen.

Ein ischämischer Schlaganfall tritt auf, wenn Blutgerinnsel sich in Hirngefäßen festsetzen und den Blutfluss im Gehirn und somit dessen Sauerstoffversorgung unterbrechen. Eine mögliche Maßnahme in diesem Fall ist die mechanische Thrombektomie, ein minimalinvasiver Eingriff, bei dem das Gefäß mit einem speziellen Katheter wieder geöffnet wird. Ob die mechanische Thrombektomie für die betroffene Person jedoch die vielversprechendste Option darstellt, hängt von einer Vielzahl individueller Faktoren ab. Um Ärzt:innen bei dieser zeitkritischen Entscheidung zu unterstützen, wollen die Forschenden in Liberate AI ein KI-Modell mit medizinischen Daten aus dem Deutschen Schlaganfall-Register sowie den zugehörigen MRT- und CT-Aufnahmen aus verschiedenen deutschen Krankenhäusern trainieren. Hierfür nutzen sie Swarm Learning, eine vom DZNE in Kooperation mit Hewlett Packard Enterprise entwickelte KI-Technologie. Swarm Learning ermöglicht es der KI, dezentral zu lernen: Sie reist virtuell zu allen Datenquellen im Netzwerk und sammelt dort Wissen ein, ohne dass die Daten selbst die Standorte verlassen, an denen sie gespeichert sind.

Auf dem Weg zur Glasbox: Erklärbarkeit ist entscheidend

In Liberate AI werden zudem technologische Herausforderungen adressiert, die über das eigentliche Training des KI-Modells hinausgehen. Die erste große dieser Herausforderungen betrifft die Erklärbarkeit des KI-Modells. Im Gegensatz zu Deep-Learning-Anwendungen, die meist wie eine Black Box funktionieren, muss das KI-Modell in Liberate AI seine Entscheidungsfindung für die behandelnden Ärzt:innen transparent nachvollziehbar machen. Prof. Dr. Jilles Vreeken, Experte für vertrauenswürdige Informationsverarbeitung am CISPA, erklärt: „Wir wollen eine Glasbox-KI entwickeln, die genauso gute Vorhersagen trifft wie eine Black-Box-KI. Denn wenn man Mediziner:in ist und die KI sagt ‚Ja‘ oder ‚Nein‘, dann ist die erste Frage, die man stellt: ‚Warum sollte ich dir vertrauen?‘. Das bedeutet, dass wir erklärbare KI einsetzen müssen – also den KI-Forschungszweig, in dem wir KI-Modelle entwickeln, bei denen wir nachvollziehen können, auf Grundlage welcher Beweise sie ihre Aussagen treffen. Das ist die Art von KI, die Expert:innen wirklich unterstützen kann, denn Mediziner:innen sind dann in der Lage zu unterscheiden, ob die Vorhersage auf zufälligen Beweisen oder auf echten Biomarkern beruht.“

Vreeken und seine Forschungsgruppe haben es sich zum Ziel gesetzt, ein transparentes KI-Modell zu entwickeln. Im Kontext von Swarm Learning bringt Erklärbarkeit jedoch besondere technologische Herausforderungen mit sich. „Wir müssen bedenken, dass wir zwar diese Glasbox-KI entwickeln können, sie muss aber immer noch in der Lage sein, in einer Swarm-Learning-Umgebung

zu lernen und genauso zuverlässige Vorhersagen zu treffen wie eine Black-Box-KI. Es ist nicht trivial, das möglich zu machen“, so der CISPA-Forscher. Im Zuge des Projekts müssen die Forschenden daher ein Gleichgewicht finden zwischen dem Transparenzgrad des KI-Modells und seiner Fähigkeit, erfolgreich am Swarm Learning teilzunehmen.

Auf der Suche nach Subpopulationen und kausalen Schlussfolgerungen

Die zweite große Herausforderung, mit der sich die CISPA-Forschenden befassen, betrifft die Identifizierung solcher Patientengruppen, die hinsichtlich ihrer langfristigen Lebensqualität positiv oder negativ auf eine mechanische Thrombektomie reagieren. Im Idealfall wird das KI-Modell in der Lage sein, diese statistischen Subgruppen automatisch anhand bestimmter Muster zu identifizieren, die es aus den gesammelten medizinischen Daten extrahiert. „Die Frage ist: Können wir eine Glasbox-KI entwickeln, die die Bedingungen erkennen kann, unter denen Menschen ein außergewöhnliches Überlebensverhalten zeigen? Zum Beispiel könnte das von der Größe des Blutgerinnsels, hohem oder niedrigem Blutdruck, genetischen Faktoren oder der Einnahme von Blutverdünnern abhängen. Man kann sich verschiedene Bedingungen vorstellen, die auf einige, aber nicht auf alle Patient:innen zutreffen“, erklärt Vreeken. Diese Subgruppen, betont er, können selbst dann noch identifiziert werden, wenn sich das Training eines erklärbaren Glasbox-Modells im Swarm Learning als unmöglich herausstellen sollte. „Das Schöne an unserer Glasbox-KI ist, dass wir sie zusätzlich zu einer Black-Box-KI nutzen können. Wir können nämlich fragen: ‚Für welche Menschen trifft die Black-Box-KI besonders zuverlässige Vorhersagen?‘ Selbst wenn wir also letzten Endes eine Black-Box-KI verwenden, weil sie akkurater ist als jedes transparente Modell, das wir entwickeln können, sind wir immer noch in der Lage, die Subgruppen zu bestimmen, für die wir sie befragen sollten oder nicht.“

Liberate AI: Fachexpertise und Maschinelles Lernen verbinden

Letztendlich möchten die CISPA-Forschenden ein transparentes KI-System entwickeln, das kausale Garantien für seine Vorhersagen geben kann. Wenn es beispielsweise vorhersagen sollte, dass Bluthochdruck die Wirksamkeit der Behandlung verringert, soll es auch die Gründe dafür nennen können. „Das ist sehr schwierig umzusetzen“, erklärt Vreeken, „denn man braucht eine randomisierte Kontrollstudie, um festzustellen, ob Bluthochdruck tat-sächlich der alleinige Faktor ist oder nur ein Störfaktor – also etwas, das relevant erscheint, es aber nicht ist. Die ultimative KI, die wir entwickeln möchten, ist also eine Glasbox-KI, die sagen kann: ‚Basierend auf allen verfügbaren Schlaganfalldaten gibt es einen klaren Unter-schied zwischen ansonsten vergleichbaren Patient:innen, der sich allein durch den Blut-druck erklären lässt.‘“

Selbst wenn sich die dreifache Herausforderung – Erklärbarkeit, Identifizierung von Subgruppen und kausale Garantien – am Ende als zu ambitioniert herausstellen sollte, ist Vreeken überzeugt, dass Liberate AI einen bedeutenden Beitrag zur Anwendbarkeit von KI in der Medizin leisten wird. Besonders die Interdisziplinarität des Projektteams eröffnet neue Möglichkeiten für die Behandlung akuter Schlaganfälle, wie er hervorhebt: „Wenn man Fachleute fragt, was sie wollen, dann wollen sie eine bessere Maschine X. Vielleicht brauchen sie aber etwas ganz anderes, von dem sie gar nicht wissen, dass es möglich ist. Das gegenteilige Problem ist, dass Informatiker:innen oft neue Maschinen entwickeln, von denen die Fachleute vielleicht sagen: ‚Das löst ein Problem, das wir gar nicht haben.‘ Ich bin sehr froh, dass wir in diesem Projekt eine hervorragende Konstellation von Menschen mit Informatikexpertise, Menschen mit rein medizinischer Expertise und Menschen dazwischen haben. In Liberate AI werden wir keine Maschine entwickeln, auf die niemand wartet – sondern eine Maschine, von der die Menschen überhaupt nicht wissen, dass sie sie brauchen.“

Dieses Forschungsprojekt wird unterstützt durch den Impuls- und Vernetzungsfond von Helmholtz.

Über Liberate AI

Liberate AI ist ein gemeinsames Forschungsprojekt von Forschenden des Deutschen Zentrums für Neurodegenerative Erkrankungen (DZNE), der Abteilung für Vaskuläre Neurologie am Universitätsklinikum Bonn (UKB) und des CISPA Helmholtz-Zentrums für Informationssicherheit. In Liberate AI soll ein auf künstlicher Intelligenz (KI) basierendes Computermodell entwickelt werden, das Ärzt:innen bei der Behandlung von akuten Schlaganfällen unterstützt. Als digitales Assistenzsystem soll das Modell den langfristigen Behandlungserfolg nach einer minimalinvasiven Therapie (mechanische Thrombektomie) sowie mögliche Komplikationen vorhersagen und so bei der Entscheidung über die bestmögliche Therapie helfen. In das Training des KI-Modells sollen zentrale Registerdaten sowie lokal vorhandene Aufnahmen des Gehirns einfließen. Liberate AI steht unter der Leitung von Prof. Dr. Joachim Schultze am DZNE und wird von der Helmholtz-Gemeinschaft mit 250.000 Euro gefördert.

Über CISPA

Das CISPA Helmholtz-Zentrum für Informationssicherheit ist eine nationale Großforschungseinrichtung in der Helmholtz-Gemeinschaft. Es erforscht Informationssicherheit in all ihren Facetten, um die drängenden großen Herausforderungen der Cybersicherheit und der vertrauenswürdigen Künstlichen Intelligenz, mit denen unsere Gesellschaft im digitalen Zeitalter konfrontiert ist, umfassend und ganzheitlich anzugehen. Das Zentrum nimmt eine weltweite Führungsrolle im Bereich der Cybersicherheit ein, indem es exzellente und oft auch disruptive Grundlagenforschung mit innovativer angewandter Forschung, Technologietransfer und gesellschaftlichem Diskurs verbindet. Thematisch deckt es das gesamte Spektrum von der Theorie bis zur empirischen Forschung ab. International ist es als Ausbildungsstätte für die nächste Generation von Cybersicherheits-Expert:innen sowie wissenschaftlichen Führungskräften in diesem Bereich anerkannt.