

## Sprachmodell UroBot beantwortet Fachfragen genauer als Urologen

**Wissenschaftler des Deutschen Krebsforschungszentrums (DKFZ) haben gemeinsam mit Ärzten der Urologischen Universitätsklinik Mannheim einen auf künstlicher Intelligenz basierenden Chatbot entwickelt und erfolgreich getestet. "UroBot" konnte Fragen der Facharztprüfung Urologie mit hoher Genauigkeit beantworten und übertrifft dabei sowohl andere Sprachmodelle als auch die Genauigkeit erfahrener Urologen. Das Modell begründet seine Antworten detailliert anhand der Leitlinien.**

Mit den Fortschritten der personalisierten Onkologie werden urologische Leitlinien zunehmend komplexer. Sei es im Tumorboard, auf Station oder in der Praxis: Ein präzises Zweitmeinungssystem für medizinische Entscheidungen in der Urologie könnte Ärzte bei der evidenzbasierten und personalisierten Versorgung unterstützen, insbesondere bei Zeit- oder Kapazitätsengpässen. Große Sprachmodelle (LLMs) wie GPT-4 haben zwar das Potenzial, medizinisches Fachwissen abzurufen und ohne zusätzliche Schulung komplexe medizinische Fragen zu beantworten. Allerdings ist ihre Anwendbarkeit in der klinischen Praxis oft aufgrund veralteter Trainingsdaten und mangelnder Erklärbarkeit eingeschränkt. Um diese Hürden zu überwinden, hat ein Team um Titus Brinker vom DKFZ "UroBot", entwickelt, einen spezialisierten Chatbot für die Urologie, der durch die aktuellen Richtlinien der Europäischen Gesellschaft für Urologie ergänzt wurde.

UroBot basiert auf dem leistungsstärksten Sprachmodell GPT-4o von OpenAI. Es nutzt eine durch das Team angepasste Methode zur abrufunterstützten Generierung (Retrieval Augmented Generation, RAG), das in der Lage ist, auf die individuelle Frage hin relevante Informationen aus hunderten Dokumenten zielgerichtet abzurufen, um präzise und erklärbare Antworten zu liefern. Das modifizierte Modell wurde an 200 Fachfragen des European Board of Urology getestet und in mehreren Durchläufen bewertet.

UroBot-4o beantwortete Fragen zur Facharztprüfung zu 88,4 Prozent korrekt und übertraf damit das aktuellste Modell GPT-4o um 10,8 Prozentpunkte. Damit stellt UroBot nicht nur andere Sprachmodelle in den Schatten, sondern übertrifft auch die durchschnittliche Leistung von Urologen bei der Facharztprüfung, die in der Literatur mit 68,7 Prozent angegeben wird. Zudem zeigt UroBot eine sehr hohe Zuverlässigkeit sowie Konsistenz in den Antworten.

UroBots Antworten sind durch klinische Experten verifizierbar, da die entscheidenden Quellen und Textabschnitte durch die Software benannt werden: „Die Studie zeigt, welches Potenzial in der Verbindung von großen Sprachmodellen mit evidenzbasierten Leitlinien steckt, um die Leistung in spezialisierten medizinischen Bereichen zu verbessern. Die Verifizierbarkeit und die gleichzeitig sehr hohe Genauigkeit machen UroBot zu einem vielversprechenden Assistenzsystem für die Patientenversorgung. Die Nutzung von nachvollziehbaren Sprachmodellen wie UroBot wird in den nächsten Jahren eine enorme Bedeutung in der Patientenversorgung bekommen und helfen, leitliniengerechte Versorgung auch bei zunehmend hoher Komplexität von Therapieentscheidungen flächendeckend sicherzustellen“, so Brinker.

Das Forschungsteam hat den Code und die Anleitung zur Nutzung von UroBot veröffentlicht, um künftige Weiterentwicklungen in der Urologie, aber auch weiteren medizinischen Fachbereichen zu

ermöglichen.

Martin J. Hetz, Nicolas Carl, Sarah Haggemüller, Christoph Wies, Jakob Nikolas Kather, Maurice Stephan Michel, Frederik Wessels, Titus J. Brinker: Superhuman Performance on Urology Board Questions Using an Explainable Language Model Enhanced with European Association of Urology Guidelines

ESMO Real World Data and Digital Oncology 2024,

DOI: <https://doi.org/10.1016/j.esmorw.2024.100078>