

Allgemeines

Ich arbeite gerade an meiner Bachelorarbeit, in der ich evaluiere, wie gut KI-Modelle (speziell Large Language Modelle) darin sind, relevante Informationen aus medizinischen Richtlinien zu extrahieren, die später medizinisches Personal unterstützen soll, schneller Überblick über den Zustand des Patienten zu erhalten und Diagnoseentscheidungen zu treffen.

Diese Informationen sind zum einen Observations, also so etwas wie Blutdruck, Körpertemperatur, Reaktionsvermögen usw., mit unterschiedlichen Wertebereichen und Grenzen. Diesen wird ein Relevanz-Score zugeordnet, der beschreiben soll, wie wichtig diese Observation für das zuständige medizinische Personal ist, abhängig davon, wie sehr sie vom "Normalwert" abweicht.

Zum anderen gibt es auch Conditions, welche medizinischen Diagnosen wie Hypertension, Akuter Appendizitis usw. entsprechen. Diesen werden dann Observations zugeordnet, und wieder ein Relevanz-Score vergeben, wie sehr diese Observation auf die Condition/Diagnose hinweist.

Um nun Evaluieren zu können, wie gut die LLMs darin sind, die relevanten Informationen zu extrahieren, brauche ich eine Ground Truth, also Daten, die von medizinischem Personal extrahiert wurden. Um es nicht zu umständlich zu machen, habe ich diese Extraktion schon durchgeführt, und würde nun gerne Ihre Hilfe in Anspruch nehmen, zu kontrollieren, dass diese extrahierten Informationen richtig/relevant sind. Dabei sind die exakten Relevanzwerte in meinen gegebenen Informationen nicht wichtig zu kontrollieren, da diese in der Evaluation nicht berücksichtigt werden.

Falls Ihnen Fehler auffallen oder Sie Änderungen sehen, die näher an den Daten ist, auf die sich medizinisches Personal fokussieren würde, dann schreiben sie dies gerne auf/ändern die Tabelle

Vielen Dank im Voraus, ich schätze Ihre Unterstützung sehr.