

## Über dianovis Software:

In der Softwarearchitektur von dianovi wird eine RAG (Retrieval-Augmented Generation)-Architektur genutzt, die auf Basis von Patientenfällen relevante Abschnitte aus klinischen Leitlinien identifiziert. Mithilfe dieser Architektur liefert die Software präzise Antworten, die optimal auf die individuellen Patientenfälle abgestimmt sind und den medizinischen Entscheidungsprozess gezielt unterstützen.

---

## Masterarbeit/Bachelorarbeit 1:

### „Optimierung der Wissensdatenbank für KI-basierte Leitliniennutzung im klinischen Kontext“

#### Thema:

Ziel dieser Arbeit ist es, die Wissensdatenbank eines klinischen KI-Systems durch gezieltes Preprocessing und die Verdichtung relevanter Textinhalte zu optimieren. Der Fokus liegt auf der automatisierten Extraktion, Vorverarbeitung und komprimierten Zusammenfassung klinischer Leitlinien, um die Effizienz und Präzision der Informationsbereitstellung für ärztliche Nutzer zu verbessern.

#### Schwerpunkte:

- **Entwicklung eines Preprocessing-Ansatzes** zur effizienten Verarbeitung medizinischer Leitlinien. Hierzu gehört die Entfernung von Verzeichnissen und irrelevanten Textpassagen sowie die Entwicklung eines Verfahrens zur Integration und Darstellung ergänzender Leitlinieninhalte, wie Abbildungen und Beschreibungen, unter Einsatz neuronaler Netze oder Transformermodelle zur Textextraktion.
- **Verdichtung relevanter Abschnitte** mithilfe von LLMs, um komprimierte und kontextbezogene Zusammenfassungen zu generieren, die anschließend von Experten validiert werden.
  - **Proposition Indexing (Quelle)**: Erstellung hochinformativer Chunks, um Leitlinienaussagen effizienter abzubilden.
  - **RAPTOR (Quelle)**: Clusterung und Zusammenfassung komplexer Chunks über Leitlinien hinweg.
- **Reduzierung der benötigten Chunk-Anzahl**: Untersuchung, ob die verkürzte Darstellung der Leitlinien die Anzahl der Chunks minimieren und somit eine präzisere und effizientere Verarbeitung durch das KI-System ermöglichen kann.

#### Forschungsfrage:

Wie kann die Wissensdatenbank eines klinischen KI-Systems durch automatisierte Extraktion und Verdichtung klinischer Leitlinien optimiert werden, um die Effizienz und Genauigkeit der Informationsbereitstellung für medizinische Fachkräfte zu steigern?

---

## Masterarbeit/Bachelorarbeit 2:

## „Optimierung der Retrieval-Struktur für semantisches Retrieval in klinischen Leitliniendatenbanken“

### Thema:

Diese Arbeit zielt darauf ab, die Retrieval-Struktur eines klinischen KI-Systems zu optimieren, um eine präzisere Identifikation und Auswahl relevanter Leitlinieninformationen zu gewährleisten. Der Fokus liegt auf der Evaluierung und Implementierung fortschrittlicher Retrieval-Techniken, die die Trefferquote steigern und Fehlerquoten minimieren.

### Schwerpunkte:

- **Analyse und Auswahl geeigneter Embedding-Modelle** zur Optimierung der Verarbeitung von Chunks mit mindestens 1.000 Tokens, um inhaltlich umfangreiche und präzise Ergebnisse zu erzielen.
- **Implementierung und Vergleich moderner Retrieval-Techniken, wie z.B.:**
  - **ColBERT** ([Quelle](#)): Tokenbasierte Zerlegung, die die Kontexteffizienz durch maximale Übereinstimmung verbessert.
  - **Multi-Query** ([Quelle](#)): Anwendung mehrfacher Abfragen zur Minimierung von Doppeldeutigkeiten durch Umformulierungen.
- **Optimierung der Retrieval-Struktur und Vergleich verschiedener Techniken:** Steigerung der Relevanz und Genauigkeit der Datenbankabfragen, um die Antwortgenauigkeit für klinische Anfragen zu maximieren und Fehlertoleranzen zu minimieren.

### Forschungsfrage:

Wie kann die Retrieval-Struktur eines klinischen KI-Systems angepasst werden, um die Relevanz und Genauigkeit von Leitlinienabfragen zu maximieren und die Fehlerquote im klinischen Kontext zu minimieren?