

Stefan Schulz<sup>1</sup>, Philipp Daumke<sup>1</sup>, Torsten Koller<sup>1</sup>, Martin Boeker<sup>2</sup>

<sup>1</sup>Averbis GmbH, Freiburg, <sup>2</sup>Institut für KI und Informatik in der Medizin – Technische Universität München

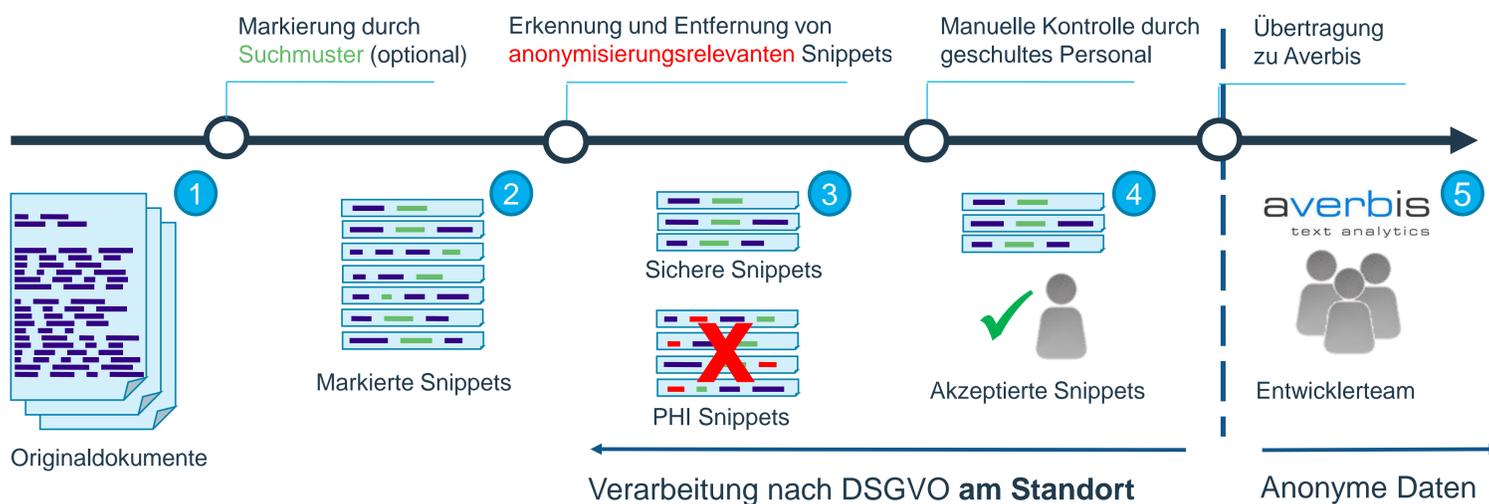
## Die Performanz eines Textanalyse- (NLP-) Systems hängt davon ab, wie gut es mit authentischen Texten trainiert wird

- Akquisition lexikalischer Einheiten
- Manuelles Erstellen von Regeln
- Maschinelles Lernen von Modellen auf der Basis annotierter Korpusdaten

## An die Nutzung klinischer Dokumente als Ressource – ohne Zustimmung der Patient\*innen – sind höchste Sicherheitsanforderungen zu stellen

- Dokumententypen: Arztbriefe, Befundtexte
- De-Identifizierung: automatisch mit manueller Kontrolle
- Unterteilung in kleine Einheiten ("Snippets") – maximal einzelne Sätze

- Ausgangspunkt: MIRACUM D2.1: Nachweis des flächendeckenden **Einsatzes von NLP-Technologie** in den DIZ → Publikation
- Plan: In den DIZ erzeugte große Mengen **anonymisierter Textfragmente** (Snippets) werden zum Industriepartner **Averbis** übertragen und dort in den Produktionsprozess übernommen
- Beteiligung DIFUTURE: Erstellung eines **Studienprotokolls**. Endpunkt: Verbesserte Erkennung von Diagnosen und Medikationen nach Trainieren des Systems mit Echtdateien
- Pilotierung durch DIZ des Uniklinikums Freiburg nach **positivem Ethikvotum** und Zustimmung der Datenschutzbeauftragten



- Über die Art der Dokumente (Fachabteilung, Dokumententyp, etc.) entscheidet jeder Standort eigenständig
- Möglich sind alle Dokumente, die klinisch relevante Informationen (in unstrukturierter Form) enthalten
- Richtwert: Extraktion von Snippets aus ca. 10.000 Dokumenten von individuellen Patienten je Standort.

- Zerlegung von Arztbriefen in disjunkte Sätze mittels Satzzerlegung von Averbis Health Discovery
- Optional: Limitierung der Wortanzahl → Snippets
- Filterung der Snippets nach Diagnosen und Medikationen nach Annotation durch Health Discovery Diagnose- und Medikationserkennung

- Erkennung von PHI-relevanten Entitäten (Protected Health Information) Datum, Alter, ID, Ort, Name, Adresse, Beruf innerhalb Snippets\*
- Verwendung der automatisierte PHI-Erkennung der Averbis Health Discovery
- Aussortierung aller Snippets in denen mindestens eine PHI-relevante Entität gefunden wurde.

- Durchsicht der verbleibenden Snippets durch geschulte Person unter Aufsicht
- Ggf. Aussortieren kritischer Snippets
- Entscheidung, wer die Durchsicht übernehmen darf, durch den jeweiligen Standort getroffen (Datenschutzstelle)
- Finale Freigabe der anonymisierten Snippets

- Finale Freigabe seitens des DIZ
- Übertragung zu Averbis über einen sicheren Kanal
- Verwendung der Daten zur Weiterentwicklung der Softwarewerkzeuge
- Manuelle Erstellung eines Benchmarks an einer Untermenge der Snippets
- Nach Weiterentwicklung der Werkzeuge: Validierung an Benchmark

## Erste Ergebnisse aus der Pilotstudie mit Daten aus dem DIZ des Universitätsklinikums Freiburg

- Durchsicht von ca. 13.000 Snippets
- Kein einziger kompletter Namen (Vorname + Nachname) bei manueller Durchsicht
- Einige Vornamen von Kindern nicht anonymisiert („der kleine Max kam in Begleitung der Mutter“)
- Nach Personen der Medizingeschichte benannte Krankenstationen nicht anonymisiert („Station Eisenbart“)
- Wenige Namen von Krankenhäusern oder Pflegeheimen („Pflegeheim St. Dominikus“)
- Kombinationen von möglicherweise identifizierenden Namenszusätzen um anonymisierten Namen herum "Prof. Dr. mult. Xxxxxxx Xxxxx" MBA, Ärztl. Direktor

## Ausrollen an anderen Standorten von MIRACUM und DIFUTURE

- Studienbeschreibung der Annotationsstudie auf Confluence
  - Ebenso Dokumentation (Datenschutz, Ethik) am Standort Freiburg
  - Schnellere Freigabe aufgrund vorhandener Dokumentation und durchgeführter Pilotstudie zu erwarten
  - Seitens DIZ: Vorhandensein Averbis Health Discovery V6
  - Bereitstellung klinischer Dokumente nach Wahl im DIZ
  - Installation eines Skripts durch Averbis-Mitarbeiter
  - Snippet-Extraktion, Filterung durch Skript,
  - Freigabe und Übermittlung nach Vorgabe des Standorts
- Weitere Optimierung der Namenserkennung in der Averbis Health Discovery derzeit in Arbeit**

**Danksagung: Mitarbeiter\*innen von IMBI und DIZ Freiburg für die Bereitstellung von Dokumenten, aktive Unterstützung und insbesondere für die Schnittstellenfunktion zu Datenschutz und Ethikkommission**

**Kontakt:**  
Stefan Schulz, Averbis GmbH  
Salzstraße 15, 79098 Freiburg im Breisgau  
stefan.schulz@averbis.com