

# Evaluation Speech-To-Text-Algorithmen in der Dokumentation im Rettungsdiensteinsatz

Studiengang: BSc in Medizininformatik  
Vertiefung: Advanced Data Processing  
Betreuer: Prof. Dr. Murat Sariyar  
Experte: Martin Arnold (AMCons GmbH)

Dokumentationen dienen im Gesundheitswesen als wichtiges Hilfsmittel zum Wissenstransfer und -erhalt von Behandlungsinformationen. Automatisierte oder digital unterstützte Prozesse fliessen schrittweise in diese Dokumentationsarbeiten ein. In dieser Arbeit wird genauer beleuchtet, was bei automatischen Transkriptionen mittels Speech-To-Text (STT)-Algorithmen im Rettungsdiensteinsatz speziell beachtet werden muss, und wie dies implementiert werden kann.

## Ausgangslage

Mit der ansteigenden Digitalisierung im Gesundheitswesen bieten sich immer mehr die Möglichkeiten, Dokumentationsprozesse mittels digitalen Hilfsmitteln zu vereinfachen. Der Einsatz Künstlicher Intelligenz (KI) bietet viele Vorteile, um Prozesse zu automatisieren oder zu ergänzen. Wie diese spezifisch im Rettungsdiensteinsatz etabliert werden können, ist jedoch noch zu erforschen. Innerhalb von hochkritischen Situationen, wie sie im Rettungsdiensteinsatz vorkommen, muss der Erhalt von Informationen in der Dokumentation präzise und effizient erfasst werden. Es gilt herauszufinden, wie ein solcher Prozess im Zusammenhang mit Tonaufnahme und STT-Algorithmus genau aussehen muss, und wie dieser evaluiert werden kann.

## Methodik

Zur systematischen Evaluierung von STT-Technologien für den Rettungsdiensteinsatz wurde wie folgt vorgegangen:

## Literatur- und Marktanalyse

Eine Recherche zu bestehenden STT-Algorithmen sowie KI-basierten Auswertungsmethoden diente der Auswahl der zu evaluierenden Technologien und geeigneter Metriken.

## Audioerzeugung und Simulation

Damit diese Technologien korrekt geprüft werden konnten, wurden zusätzlich Text-To-Speech-Technologien untersucht, bei denen aus Texten möglichst realitätsnahe Audioaufnahmen generiert werden konnten. Hierbei wurden den Audiodateien variable Hintergrundgeräusche (z. B. Lärm, Umgebungsgeräusche) hinzugefügt, um verschiedene Einsatzszenarien zu simulieren.

## Evaluations-Pipeline

Die generierten Audiodateien wurden durch die ausgewählten STT-Technologien transkribiert. Die

Transkriptionen wurden darauffolgend anhand der definierten Metriken bewertet.

Die gesamte Pipeline wurde in Python implementiert, bestehend aus Jupyter-Notebooks und Standalone-Skripten, die auf einer GitLab-Instanz der Berner Fachhochschule (BFH) gesichert wurden. Dabei wurde auf Erweiterbarkeit und einfache Integration neuer Technologien geachtet.

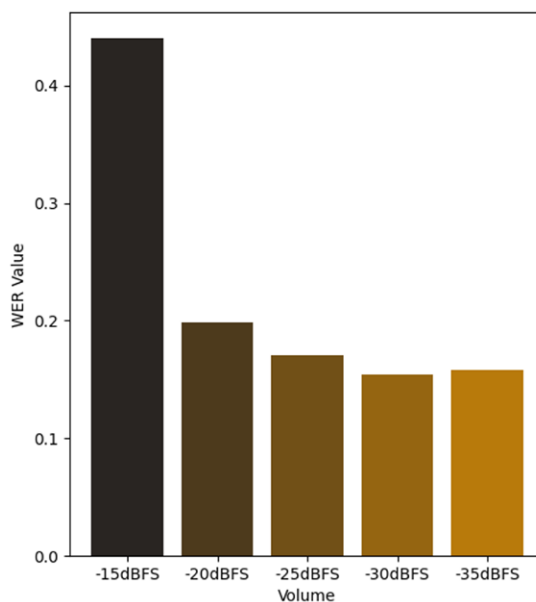
## Ergebnisse

Folgende Technologien wurden während des Projekts geprüft: Whisper von OpenAI, Recapp von Recapp IT AG und Vosk von AlphaCephei

Zur Bewertung der transkribierten Texte wurden folgende Metriken definiert: Wortfehlerrate (WER), Satzfehlerrate (SER), Übersetzungs-/ Transkriptionsqualität (BLEU), Textähnlichkeit anhand von zwei Vektoren (Cosine Similarity) und Begriffserkennung mittels F1-Score, Precision und Recall.



Nikola Stanic



Wortfehlerrate (WER) aller Transkriptionen mit Whisper (Modell «Large») im Umfeld «drinnen gedrängt»