

Compliant Text Normalization for Small and Medium-sized Enterprises — A Work-in-Progress Paper

Regina Stodden & Philipp Cimiano — Uni Bielefeld
regina.stodden@uni-bielefeld.de
Projekt: LLM4KMU

KI in KMU Workshop – Potsdam

September 16th, 2025

- LLMs können kleine und mittelständische Unternehmen (KMUs) in vielen Bereichen unterstützen^{1,2}, z.B.
 - Interaktionen mit Kunden,
 - Extraktion von Informationen,
 - Unterstützung bei Programmierung,
 - (Vor-)Erstellung von (Produkt-)Handbüchern, oder
 - Erleichterung von Dokumentation

¹ Chkurbene, Z. et al. (2024): Large Language Models (LLM) in Industry. <https://doi.org/10.1109/HONET63146.2024.10822885>

² Olujimi & Ade-Ibijola (2023): NLP techniques for automating responses to customer queries. <https://doi.org/10.1007/s44163-023-00065-5>

- LLMs können kleine und mittelständische Unternehmen (KMUs) in vielen Bereichen unterstützen^{1,2}, z.B.
 - Interaktionen mit Kunden,
 - Extraktion von Informationen,
 - Unterstützung bei Programmierung,
 - (Vor-)Erstellung von (Produkt-)Handbüchern, oder
 - **Erleichterung von Dokumentation**

¹ Chkurbene, Z. et al. (2024): Large Language Models (LLM) in Industry. <https://doi.org/10.1109/HONET63146.2024.10822885>

² Olujimi & Ade-Ibijola (2023): NLP techniques for automating responses to customer queries. <https://doi.org/10.1007/s44163-023-00065-5>

Dokumentation in KMUs

- Dokumentation spielt eine wichtige Rolle in vielen KMU-Sparten, z.B. mit Kundenbetreuung oder Pflegearbeit.
- Unternehmen fordern oft eine Dokumentation mit hoher Qualität, um Kunden, Angehörige oder Förderorganisationen die Aufträge, Einsätze und Dienstleistungen nachweisen zu können.
- Dokumentation ist notwendig zur Beantragung von Erstattungen oder Aufwandsentschädigungen.¹



¹ Maier et al. (2020): Supporting Social Workers with Summarizations of Patient Trajectories extracted from Documentation.
doi.org/10.1145/3416028.3416032.

- Gute Dokumentation ist allerdings sehr zeitaufwendig.
- In Pflegeberufen wird Dokumentation sogar mit erhöhter kognitiver Belastung und Ablenkung in Verbindung gebracht.²
- Unternehmen legen oft Regeln fest, um die Dokumentation zu standardisieren und die Qualität zu sichern.
- In der Praxis unterscheidet sich der Dokumentationsstil aber oft deutlich, z.B.
 - in Länge, Qualität, Schreibstil,
 - durch erlernte Dokumentationstechniken oder
 - eigene Abwandlungen.



¹ Quiroz et al. (2019): Challenges of developing a digital scribe to reduce clinical documentation burden. doi.org/10.1038/s41746-019-0190-1

- Gute Dokumentation ist allerdings sehr zeitaufwendig.
- In Pflegeberufen wird Dokumentation sogar mit erhöhter kognitiver Belastung und Ablenkung in Verbindung gebracht.³
- Unternehmen legen oft Regeln fest, um die Dokumentation zu standardisieren und die Qualität zu sichern.
- In der Praxis unterscheidet sich der Dokumentationsstil aber oft deutlich, z.B.
 - in Länge, Qualität, Schreibstil,
 - durch erlernte Dokumentationstechniken oder
 - eigene Abwandlungen.



KI-gestützte Dokumentation als Lösungsvorschlag

¹ Quiroz et al. (2019): Challenges of developing a digital scribe to reduce clinical documentation burden. doi.org/10.1038/s41746-019-0190-1

KI-gestützte Dokumentation — regelkonforme Textnormalisierung

- Normalisierung von Texten der Dokumentation zu einheitlichen Texten, die
 - die Regeln des Unternehmens einhalten
 - und vorzeigbar sind

- Normalisierung von Texten der Dokumentation zu einheitlichen Texten, die
 - die Regeln des Unternehmens einhalten
 - und vorzeigbar sind
- Kombination aus verschiedenen Natural Language Processing Tasks:
 - Textnormalisierung,
 - Transfer eines Textstils und
 - Textkorrektur
- Änderung von informellen oder unvollständigen Notizen zu formeller, vorzeigbarer Dokumentation

- Normalisierung von Texten der Dokumentation zu einheitlichen Texten, die
 - die Regeln des Unternehmens einhalten
 - und vorzeigbar sind
- Kombination aus verschiedenen Natural Language Processing Tasks:
 - Textnormalisierung,
 - Transfer eines Textstils und
 - Textkorrektur
- Änderung von informellen oder unvollständigen Notizen zu formeller, vorzeigbarer Dokumentation
- Änderungen sind auf folgenden Ebenen möglich, z.B.:
 - Textstil, Wortwahl, oder Satzstruktur

Beispieldokumentation — Pflege

- Tägliche Fortschrittsberichte oder Pflegevermerke zu Zustand, Medikation und erbrachten Dienstleistungen pro Klient
 - Standardrichtlinie für Dokumentation: SOAP Richtlinie
 - Beschreibung des Klienten aus eigener Sicht: subjective (S) ,
 - Beschreibung des Klienten aus Sicht des Personals: objective (O) ,
 - Bewertung der Verfassung des Klienten: assessment (A) ,
 - Weiterer Plan für den Klienten: plan (P)
- ➔ Kooperation mit kleinen und mittelständischen Unternehmen aus der Pflegebranche



1) Pflegekraft formuliert Notizen vor (entweder schriftlich oder verbal)



2) KI-Modelle überprüfen die Notizen auf Einhaltung der Richtlinien (im Hintergrund)



3) KI-Modelle machen Vorschläge zur Verbesserung (simultan angezeigt zum Text) oder zeigen verbesserten Text



4) Pflegekraft überprüft die Vorschläge und Anmerkungen und stimmt ihnen zu oder lehnt sie ab



5) Nach den Änderungen kann der Text solange erneut überprüft werden bis das Ergebnis zufriedenstellend ist

subjective objective assessment plan

AB. Keine Beschwerden, spielt und unterhält sich mit anderen med ok.

(a) Exemplarische, kurze Version

subjective objective assessment plan

AB. Keine Beschwerden, spielt und unterhält sich mit anderen med ok.

(a) Exemplarische, kurze Version

Herr Peters begann seinen Tag mit einem Kartenspiel mit anderen Bewohnern. Es wurde beobachtet, dass er kontaktfreudig ist und sich an Gesprächen beteiligt. Die Medikamenteneinnahme erfolgt ohne Probleme.

(b) Exemplarische, lange Version

subjective objective assessment plan

AB. Keine Beschwerden, spielt und unterhält sich mit anderen med ok.

(a) Exemplarische, kurze Version

Herr Peters begann seinen Tag mit einem Kartenspiel mit anderen Bewohnern. Es wurde beobachtet, dass er kontaktfreudig ist und sich an Gesprächen beteiligt. Die Medikamenteneinnahme erfolgt ohne Probleme.

(b) Exemplarische, lange Version

Herr Peters hat keine Probleme erwähnt. Er begann seinen Tag mit einem Kartenspiel mit anderen Bewohnern. Ich habe beobachtet, dass er kontaktfreudig ist und sich an Gesprächen beteiligt. Die Medikamenteneinnahme erfolgt ohne Probleme. Es sind keine Änderungen der Medikation erforderlich. Keine weiteren Pläne.

(c) Normalisierte Version

Methodik

- Verwendung von kleineren open-source Modellen, wie LLama, Mistral oder BLOOMZ, um das Budget von KMU zu schonen
- Fokus auf robuste Modelle und Abschwächung von Halluzinationen
- Herangehensweise:
 - Schritt-für-Schritt Ansatz
 - Persona Ansatz
 - Human-in-the-loop Ansatz

- Die LLMs erhalten einen Prompt, in dem die Herangehensweise zur Erstellung der normalisierten Texte Schritt-für-Schritt erklärt wird
 1. unvollständige Halbsätze ausformulieren
 2. Anpassung des Schreibstils (umgangsprachlich zu formell)
 3. Regeln überprüfen und anwenden

- Die LLMs erhalten einen Prompt, in dem das erwartete Verhalten und Erfahrung präzisiert wird, z.B.
 - langjährige Arbeit in der Pflege
 - tägliche Erfahrung mit der Formulierung von Pflegevermerken
 - Einhaltung des Regelwerks ist der Person besonders wichtig

- Mit einem der vorherigen Ansätze wird ein erster Vorschlag generiert
- Eine Pflegekraft liest den Vorschlag gegen und optimiert ihn
- Anschließend werden die Änderungsvorschläge oder der verbesserte Text dem Modell zur Verfügung gestellt, um daraus zu lernen
- Alternativ könnten dem Modell Qualitätsbewertungen der generierten Texte gefüttert werden

- Automatische Metriken

- Metriken zur Messung von Bedeutungserhaltung:

- ▶ BLEU
- ▶ BERTScore
- ▶ ROUGE

- LLM as a Judge

➔ anzweifelbare Qualität

- Manuelle Evaluation

- Bedeutungserhaltung,
- Grad der Regelkonformität,
- Faktizität und Korrektheit oder
- Formalität

→ aufwendig, aber vertrauenswürdiger

Zusammenfassung

- Zusammenfassung:
 - Erste Einblicke in das Projekt zu regelkonformer Textnormalisierung
 - Regelkonforme Textnormalisierung kann KMUs im Dokumentationsprozess unterstützen
- Ausblick:
 - Aktuell Datenbeschaffung
 - Erste Ergebnisse bald zu erwarten
- Erweiterungspläne:
 - Kombination von regelkonformer Textnormalisierung mit Plausibilitäts- und Vollständigkeitscheck

Vielen Dank für die Aufmerksamkeit!

Compliant Text Normalization for Small and Medium-sized Enterprises — A Work-in-Progress Paper

Regina Stodden — Uni Bielefeld
regina.stodden@uni-bielefeld.de
Projekt: LLM4KMU

Fragen oder Anregungen? 