

## **(Arbeits-)Titel**

Evaluierung des Einflusses von Datenqualitätsproblemen auf die Robustheit von Transformer-basierten Named Entity Recognition-Modellen

## **Problemstellung**

*Bitte beschreiben Sie Ihre Problemstellung ausführlich und sorgfältig. Eine gut ausformulierte und vorher abgestimmte Beschreibung hilft Ihnen bei der tatsächlichen Bearbeitung Ihrer wissenschaftlichen Arbeit.*

Die automatische Erkennung von Entitäten in Texten (Named Entity Recognition, NER) ist ein zentraler Baustein im Bereich Natural Language Processing (NLP). Transformer-basierte Modelle wie BERT oder RoBERTa erreichen auf sauberen, gut annotierten Datensätzen beeindruckende Genauigkeiten. In der Praxis treten jedoch häufig Qualitätsprobleme in Trainings- und Eingabedaten auf, etwa orthographische Fehler, fehlerhafte Annotationen (Label-Noise) oder syntaktische Störungen. Diese Datenverschlechterungen führen oft zu drastischen Leistungseinbußen. Im Moment existiert nur eine begrenzte Übersicht, welche Arten von Datenverschlechterung die Modelle besonders beeinträchtigen und bei welchen Schwellenwerten die Modelle in ihrer Leistung drastisch abfallen.

## **Zielsetzung der Arbeit**

*Die Zielsetzung soll möglichst präzise formuliert sein und ggf. auch die Grenzen der Arbeit verdeutlichen. (z.B. Was ist nicht Ziel der Arbeit?)*

Ziel der Arbeit ist es, den Einfluss verschiedener Datenqualitätsprobleme auf Transformer-basierte NER-Modelle systematisch zu untersuchen und daraus fundierte Handlungsempfehlungen abzuleiten. Dabei wird zunächst im Rahmen eines systematischen Literature Reviews (SLR) ein umfassender Überblick über bestehende Ansätze zur Simulation und Untersuchung von Datenverschlechterungen gegeben. Es soll erfasst werden, welche Arten von Störungen (z.B. orthographische Fehler, syntaktische Fehler, semantische Veränderungen, Label-Noise) existieren, welche Verfahren zur Erzeugung dieser Störungen angewandt werden und welche Effekte in bisherigen Arbeiten beobachtet wurden. Anschließend werden relevante Verschlechterungsverfahren basierend auf der Literaturrecherche ausgewählt und implementiert.

Im Fokus steht insbesondere die Untersuchung, welche Typen von Datenqualitätsstörungen die größte Beeinträchtigung verursachen und ab welchen Schwellenwerten die Modellleistung signifikant abnimmt. Dabei sollen auch Unterschiede zwischen Modellen wie BERT, RoBERTa, XLNet, DistilBERT, CANINE herausgearbeitet werden. Ergänzend wird analysiert, inwiefern Domänenspezifika wie Schreibstil, Fachvokabular oder Textstruktur die Anfälligkeit gegenüber Störungen beeinflussen. Ziel ist es, typische Muster zu identifizieren, bei denen NER-Modelle besonders leicht beeinträchtigt werden

Die Arbeit soll nicht nur die Schwächen heutiger Systeme systematisch aufzeigen, sondern auch einen praktischen Beitrag leisten: Indem die kritischsten Fehlerquellen identifiziert werden, können zukünftige Trainings- und Datenaufbereitungsstrategien gezielter auf die robustheitsrelevanten Aspekte ausgerichtet werden. Somit trägt die Arbeit zur Verbesserung der Praxistauglichkeit von NER-Modellen bei.

Explizit nicht Ziel der Arbeit ist die Entwicklung neuer Modellarchitekturen oder die umfassende Untersuchung multilingualer oder domänenübergreifender Fragestellungen. Trainingsstrategien zur Robustheitssteigerung werden nur theoretisch und nicht empirisch behandelt.

### **Geplantes Vorgehen**

Bitte beschreiben Sie Ihre geplante Vorgehensweise detailliert. Dies hilft der wiss. Leitung sehr, die Erfolgswahrscheinlichkeit und Realisierbarkeit Ihrer Arbeit einzuschätzen und Ihnen ggf. wertvolles Feedback zu geben.

1. Einleitung
  - 1.1. Motivation und Problemstellung
  - 1.2. Zielsetzung und Forschungsfrage
  - 1.3. Abgrenzung
  - 1.4. Aufbau der Arbeit
2. Grundlagen
  - 2.1. Named Entity Recognition
  - 2.2. Transformer
  - 2.3. Evaluationsmetriken
3. Methodik
4. Systematic Literature Review
  - 4.1. Research Design
  - 4.2. Vorgehensweise
  - 4.3. Ergebnisse
5. Experimente
  - 5.1. Auswahl Datensätze
  - 5.2. Vorbereitung
  - 5.3. Entwicklung Baseline Modell
  - 5.4. Implementierung von Qualitätsminderungsverfahren
6. Evaluation
  - 6.1. Analyse der Schwellenwerte und Leistungsverluste
  - 6.2. Mustererkennung bei Fehleranfälligkeit
  - 6.3. Diskussion der Unterschiede zwischen Modellen und Fehlerarten
7. Fazit, Limitationen, Ausblick

### **Wichtigste Literatur zum Thema**

Bitte geben Sie die 3-5 wichtigsten Quellen für Ihre Problemstellung an.

**Lin, B. Y.; Gao, W., Yan, J.; Moreno, R.; Ren, X. (2021):** RockNER: A Simple Method to Create Adversarial Examples for Evaluating the Robustness of Named Entity Recognition Models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 3728–3737, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

**Devlin J., Chang M-W, Lee K., Toutanova K. (2019):** BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

**Bhadauria, D; Múnera, A. S.; Krestel, R. (2024):** The Effects of Data Quality on Named Entity Recognition. In Proceedings of the Ninth Workshop on Noisy and User-generated Text (W-NUT 2024), pages 79–88, San Ġiljan, Malta. Association for Computational Linguistics.

**Merdjanovska, E., Aynetdinov, A., & Akbik, A. (2024):** NoiseBench: Benchmarking the Impact of Real Label Noise on Named Entity Recognition. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 18182–18198, Miami, Florida, USA. Association for Computational Linguistics.

**Moradi, M.; Samwald, M. (2021):** Evaluating the Robustness of Neural Language Models to Input Perturbations. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 1558–1570, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.