



[Prof. Dr. Jan Niehues //
**Künstliche Intelligenz für
Sprachtechnologien**]

Jan Niehues ist Professor für Künstliche Intelligenz für Sprachtechnologien (AI4LT) am Institut für Anthropomatik und Robotik (IAR). Sein Forschungsschwerpunkt ist die Entwicklung von Sprachtechnologien mittels künstlicher Intelligenz, im Besonderen Methoden des „Deep Learnings“. Die Forschung umfasst die maschinelle Übersetzung, die Übersetzung gesprochener Sprache, die automatische Spracherkennung sowie die Dialogmodellierung.

92

Jan Niehues studierte Informatik an der Universität Karlsruhe (TH), dem heutigen Karlsruher Institut für Technologie (KIT) und promovierte 2014 zum Thema „Adaptation in Machine Translation“. Nach seiner Postdoc Zeit am KIT mit Aufhalten bei LIMSI-CNRS in Frankreich und an der Carnegie Mellon University, Pittsburgh, PA, USA war er von 2019-2022 Assistant Professor an der Universität Maastricht. Seit 2022 ist Jan Niehues Professor am KIT und leitet die Gruppe „Künstliche Intelligenz für Sprachtechnologien“.

Jan Niehues hat an dem Erfolg zahlreicher EU-Projekte (Elitr, QT21, EU-BRIDGE) mitgewirkt sowie mehrere Industriekooperationen (Facebook, SAP) koordiniert. Er verfasste bislang über 100 wissenschaftliche Arbeiten. Darüber hinaus ist er einer der Organisatoren der International Conference on Spoken Language Translation (IWSLT) sowie Mitglied in den Programmkomitees von zahlreichen internationalen Konferenzen im Bereich der Verarbeitung natürlicher Sprache.

// **Überblick**

Die Forschungsgruppe „Künstliche Intelligenz für Sprachtechnologien“ am Institut für Anthropomatik und Robotik (IAR) entwickelt Sprachtechnologien, die mittels künstlicher Intelligenz eine natürliche Kommunikation zwischen Mensch und Maschine ermöglichen sowie die Kommunikation zwischen Menschen verbessern. Die Forschung umfasst die maschinelle Übersetzung, die Übersetzung gesprochener Sprache, die automatische Spracherkennung sowie die Dialogmodellierung.

// **Einblicke in die Forschung**

Im täglichen Leben profitieren wir heutzutage schon permanent von Technologie, die die Verarbeitung von natürlicher Sprache verwendet. Wir benutzen z. B. Tools zur automatischen Übersetzung von Text oder Sprache, wenn wir uns im Ausland aufhalten oder benutzen digitale Assistenten um unsere Umgebung zu steuern.

Am AI4LT entwickeln wir Methoden, um den Nutzen dieser Technologie zu erweitern. Dabei erforschen und adaptieren wir Methoden der künstlichen Intelli-

genz, im Besonderen des „Deep Learnings“. Ein Fokus liegt dabei auf der Entwicklung von Methoden, die weniger Daten benötigen. Wir untersuchen zum Beispiel Techniken, um Gemeinsamkeiten zwischen Sprachen zu modellieren. Damit möchten wir nicht mehr nur zwischen den meist verwendeten Sprachen der Welt übersetzen, sondern auch Übersetzung von und nach weniger verbreiteten Sprachen ermöglichen.

Außerdem interessieren wir uns für Methoden, die Systeme der Verarbeitung natürlicher Sprache effizient an die jeweiligen Bedürfnisse der Benutzer und Anwendungen anpassen.

// Projekte

- „Computational Language Documentation by 2015“: In dem DFG-geförderten Projekt entwickeln wir automatische Methoden zur Unterstützung der Dokumentation von Sprachen, die vom Aussterben bedroht sind.
- „Kontext MT“: In Zusammenarbeit mit SAP entwickeln wir Methoden, die kontextuelle Informationen in ein Übersetzungssystem integrieren. Damit möchten wir die Übersetzung von Benutzerschnittstellen verbessern.

// Ausgewählte Publikationen

Liu, Danni, Jan Niehues, James Cross, Francisco Guzmán, und Xian Li. „Improving Zero-Shot Translation by Disentangling Positional Information“. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 1259–73.

Niehues, Jan. „Continuous Learning in Neural Machine Translation using Bilingual Dictionaries“. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021). Kyiv, Ukrain, 2021.

Dinh, Tu Anh, Danni Liu, und Jan Niehues. „Tackling Data Scarcity in Speech Translation Using Zero-Shot Multilingual Machine Translation Techniques“. In ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6222–26, 2022.

Jeuris, Pedro, und Jan Niehues. „LibriS2S: A German-English Speech-to-Speech Translation Corpus“. European Language Resources Association (ELRA), 2022.

Koneru, Sai, Danni Liu, und Jan Niehues. „Unsupervised Machine Translation On Dravidian Languages“. In Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, 55–64. Kyiv.

// Mitarbeiterinnen und Mitarbeiter

Verwaltungspersonal

Silke Dannenmaier (Sekretariat)

Wissenschaftliches Personal

Danni Liu

Sai Koneru

Tu Anh Dinh

Zhaolin Li

// Website

ai4lt.anthropomatik.kit.edu/