

## Literaturzusammenfassung

### Natürliche Sprachverarbeitung in Chatbots: Ein Literaturüberblick über aktuelle Ansätze und Transformer-Technologien

*Bachelorstudium Informatik*

#### Übersicht:

Verwendete Quellen (10 Stück).....	2
Nicht-verwendete Reserve-Quellen (5 Stück).....	12

## **Verwendete Quellen (10 Stück)**

**Bauer, W., & Warschat, J. (2021). Smart Innovation durch Natural Language Processing: Mit Künstlicher Intelligenz die Wettbewerbsfähigkeit verbessern. Carl Hanser Verlag.**

**Quellen-Typ:** Artikel

**Anzahl Zitationen:** 5 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Zeigt, wie KI zur Bewältigung der Datenflut und zur Entwicklung von Innovationsstrategien eingesetzt werden kann.
- Betont die zentrale Rolle der technologischen Entwicklung und die Notwendigkeit der ständigen Verbesserung der Innovationsfähigkeit von Unternehmen.
- Stellt dar, wie aus Big Data mithilfe von KI Smart Data wird und systematische Innovationsstrategien entwickelt werden können.

### **Inhaltsübersicht:**

- Zeigt, wie KI eingesetzt und die Informationsflut bewältigt werden kann.
- Stellt die technologische Entwicklung ins Zentrum der Innovationsentwicklung.
- Navigiert durchs Dickicht der neuen Technologien.
- Liefert eine konkrete Innovations- und Technologiestrategie.
- Zeiten, in denen ein Produkt für lange Zeit erfolgreich am Markt bestehen kann, sind längst vorbei.
- Immer kürzere Lebenszyklen von Produkten, ständig steigende Kundenanforderungen oder die hohe Unsicherheit des Marktes sind nur einige Beispiele, die zeigen, wie überlebensnotwendig es für Unternehmen sein wird (und es auch bereits ist), die eigene Innovationsfähigkeit ins Zentrum der Unternehmenspolitik zu stellen und ständig zu verbessern.
- Gleichzeitig schreitet die technologische Entwicklung rasant voran, die Datenflut nimmt weiterhin extrem zu und für Unternehmen ist es schier unmöglich, hier noch den Überblick zu behalten.
- Das Werk zeigt, wie mithilfe von KI aus Big Data Smart Data werden und systematisch eine eigene Innovations- und Technologiestrategie entwickelt werden kann.
- Es veranschaulicht, wie in Zukunft neue Technologien, neue Anwendungen und Experten identifiziert werden, stellt dar, dass der Weg klar hin geht von Einzeltechnologien zu Technologiesystemen, betont die zentrale Rolle von Patenten und vieles mehr.

**Chen, G., & Schweitzer, M. (o. J.). Transformer-Modelle und ihre Anwendungen in der natürlichen Sprachverarbeitung.**

**Quellen-Typ:** Artikel

**Link:**

[https://www.wiwi.uni-siegen.de/technologiemangement/lehre/chen\\_gong\\_-\\_1471717\).pdf](https://www.wiwi.uni-siegen.de/technologiemangement/lehre/chen_gong_-_1471717).pdf)

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

#### **Relevante Kernergebnisse:**

- Die Publikation behandelt die Evolution neuronaler Netzwerkmodelle und erklärt, warum Transformer-Modelle ältere Ansätze wie RNNs ersetzen können.
- Es werden gängige vortrainierte Transformer-Modelle klassifiziert, deren Architektur, Lernaufgaben und Feinabstimmungsverfahren verglichen.
- Die Arbeit beleuchtet zukünftige Entwicklungen in der natürlichen Sprachverarbeitung und bietet einen umfassenden Überblick über die Fortschritte und Herausforderungen im Einsatz von Transformer-Technologien.

#### **Inhaltsübersicht:**

- Die Publikation "Transformer-Modelle und ihre Anwendungen in der natürlichen Sprachverarbeitung" von Chen und Schweitzer behandelt die Kerntechnologien neuronaler Netzwerkmodelle im Bereich der natürlichen Sprachverarbeitung (NLP).
- Das Kapitel beschäftigt sich mit der Evolution neuronaler Netzwerkmodelle, beginnend mit rekurrenten neuronalen Netzwerken (RNN) und ihrer Varianten, um Sequenzdaten zu verarbeiten und komplexe NLP-Probleme zu lösen.
- Die Arbeit analysiert die Struktur des Transformer-Modells und die Gründe, warum diese Technologie die vorherigen Ansätze ersetzen kann.
- Es werden gängige vortrainierte Modelle basierend auf dem Transformer-Modell klassifiziert und verglichen, wobei die Modellarchitektur, Lernaufgaben und andere Komponenten berücksichtigt werden.
- Die Publikation beleuchtet die Konzepte und Entwicklung großer Sprachmodelle, einschließlich Datensätze für Vortraining, vortrainierter Aufgaben und Feinabstimmungsverfahren.
- Die Forschung in diesem Bereich hat nicht nur die technologische Entwicklung im Bereich NLP vorangetrieben, sondern auch wertvolle Erfahrungen und Erkenntnisse für das Verständnis und den Einsatz von vortrainierten Modellen geliefert.
- Der eigentliche Zweck von vortrainierten Modellen besteht darin, dass sie mithilfe von selbstüberwachtem Lernen Sprachwissen aus unbeschriftetem Freitext extrahieren und erlernen können.
- Die verschiedenen Arten von vortrainierten Modellen zielen alle darauf ab, Sprachwissen aus dem Freitext zu lernen, wobei ihre Unterschiede in der Modellarchitektur, den Lernaufgaben und anderen Komponenten liegen.
- Die Arbeit gibt einen Ausblick auf zukünftige Entwicklungen im Bereich der natürlichen Sprachverarbeitung und fasst die gesamte Arbeit zusammen.

#### **Einsatz von Künstlicher Intelligenz zur Sprachverarbeitung**

[https://www.de.digital/DIGITAL/Redaktion/DE/Digitalisierungsindex/Publikationen/publikation-download-ki-nlp.pdf?\\_blob=publicationFile&v=3](https://www.de.digital/DIGITAL/Redaktion/DE/Digitalisierungsindex/Publikationen/publikation-download-ki-nlp.pdf?_blob=publicationFile&v=3)

**Quellen-Typ:** Artikel

**Link:**

[https://www.de.digital/DIGITAL/Redaktion/DE/Digitalisierungsindex/Publicationen/publikation-download-ki-nlp.pdf?\\_\\_blob=publicationFile&v=3](https://www.de.digital/DIGITAL/Redaktion/DE/Digitalisierungsindex/Publicationen/publikation-download-ki-nlp.pdf?__blob=publicationFile&v=3)

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

**Relevante Kernergebnisse:**

- Große Sprachmodelle, die auf Transformer-Architekturen basieren, nutzen große Textmengen und selbstüberwachte Lernalgorithmen, um natürliche Sprache zu verarbeiten und generieren.
- NLP-Anwendungen ermöglichen eine deutlich höhere Effizienz in der Marktbeobachtung und Fehlererkennung durch das Lesen und Bewerten von Tausenden Berichten in kürzester Zeit.
- Im Jahr 2021 nutzte etwa jedes zehnte Unternehmen in Deutschland KI, vor allem in Informations- und Kommunikationsdienstleistungen, Finanzdienstleistungen und technischen Dienstleistungen.

**Inhaltsübersicht:**

- Der Einsatz von Künstlicher Intelligenz (KI) zur Sprachverarbeitung, auch bekannt als Natural Language Processing (NLP), ist eine wichtige Technologie für die Automatisierung und KI-basierte Assistenz in verschiedenen Bereichen.
- NLP ermöglicht die Verarbeitung natürlicher Sprache, wodurch Maschinen in der Lage sind, Sprache zu erfassen, zu verarbeiten, zu verstehen und zu generieren.
- Große Sprachmodelle, die oft auf Transformer-Architekturen basieren, bilden den technologischen Unterbau für viele KI-Anwendungen und nutzen große Textmengen und selbstüberwachte Lernalgorithmen.
- KI-Systeme können bei kognitiven und physischen Aufgaben assistieren und Menschen mit Beeinträchtigungen die Teilhabe am Alltag und der Arbeit ermöglichen.
- Die NLP-Anwendung kann Tausende Berichte in kürzester Zeit lesen und bewerten, was eine deutlich höhere Effizienz in der Marktbeobachtung und Fehlererkennung bietet.
- Die Güte der Daten ist entscheidend für die Entscheidungsfähigkeit von KI-Systemen, wobei die Qualität der Daten oft vom Informationstransfer zwischen Service-Fachkräften und der Beschreibung von Problemen im Servicebericht abhängt.
- Im Jahr 2021 nutzte etwa jedes zehnte Unternehmen in Deutschland KI, insbesondere in den Bereichen Informations- und Kommunikationsdienstleistungen, Finanzdienstleistungen und technischen Dienstleistungen.

**Helmold, M. (2024). Chatbots und ChatGPT. In Erfolgreiche Transformation zum digitalen Champion: Wettbewerbsvorteile durch Digitalisierung und Künstliche Intelligenz (S. 111-127). Springer Fachmedien Wiesbaden.**

**Quellen-Typ:** Artikel

**Link:** [https://link.springer.com/chapter/10.1007/978-3-658-44020-6\\_13](https://link.springer.com/chapter/10.1007/978-3-658-44020-6_13)

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- ChatGPT ist ein leistungsstarkes Modell, das menschliche Sprache analysieren und kohärente Antworten generieren kann, jedoch muss die Zuverlässigkeit der Fakten kritisch überprüft werden.
- Claude von Anthropic gilt als derzeit bester AI-Chatbot, der detaillierte und nuancierte Antworten liefert und in der Interaktion mit dem Benutzer engagiert ist.
- In Europa wird ein neues Large Language Model entwickelt, das zuverlässiger, offener, transparenter und energiesparender als ChatGPT sein soll.

### **Inhaltsübersicht:**

- Moderne Chatbots sind Systeme der künstlichen Intelligenz (KI), die in der Lage sind, menschenähnliche oder menschengleiche (humanoiden) Maschinen einzusetzen.
- Die digitale Transformation und der Einsatz von KI gehören zu den wichtigsten gesellschaftlichen und wirtschaftlichen Entwicklungen unserer Zeit.
- Das Buch beschreibt Anwendungen und Konzepte der Digitalisierung und KI, sowie Praxisbeispiele in den Bereichen Supply Chain Management, Produktion, Nachhaltigkeit und Bildungswesen.
- Die Publikation "Chatbots und ChatGPT" von Marc Helmold widmet sich dem Thema der erfolgreichen Transformation zum digitalen Champion durch Digitalisierung und KI.
- ChatGPT ist ein leistungsstarkes Modell, das menschliche Sprache analysieren und kohärente Antworten generieren kann, jedoch muss die Zuverlässigkeit der Fakten kritisch überprüft werden.
- Claude von Anthropic ist derzeit der beste AI-Chatbot, der detaillierte und nuancierte Antworten liefert und in der Interaktion mit dem Benutzer engagiert ist.
- ChatGPT-4o von OpenAI bietet ebenfalls detaillierte Antworten, kann aber langsam sein und beim Abrufen von Quellen Schwierigkeiten haben.
- In Europa wird ein neues Large Language Model entwickelt, das zuverlässiger, offener, transparenter und energiesparender als ChatGPT sein soll.

### **Irie, K. (2020). Advancing neural language modeling in automatic speech recognition (Doktorarbeit, RWTH Aachen University, Germany).**

**Quellen-Typ:** Artikel

**Link:** <http://publications.rwth-aachen.de/record/789081/files/789081.pdf>

**Anzahl Zitationen:** 14 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Die neue Architektur des Transformer-Modells wird modifiziert, um den spezifischen Aufgaben der Sprachmodellierung gerecht zu werden.
- Es wird eine eingehende Vergleichsstudie zu den derzeit besten neuronalen Sprachmodellen auf Basis von RNNs durchgeführt und mit Transformer-Modellen verglichen.
- Praktische Methoden zur Anwendung der Wissensdestillation bei der Sprachmodellierung großer Vokabulare werden vorgestellt.

### **Inhaltsübersicht:**

- Die Anwendung neuronaler Sprachmodelle in der automatischen Spracherkennung ist jetzt etabliert und weit verbreitet.
- Trotz des Eindrucks eines gewissen Reifegrads wird argumentiert, dass das volle Potenzial neuronaler Sprachmodelle noch nicht ausgeschöpft wurde.
- Diese Arbeit untersucht neue Perspektiven zur Weiterentwicklung neuronaler Sprachmodelle in der automatischen Spracherkennung.
- Es wird eine eingehende Vergleichsstudie zu den derzeit besten neuronalen Sprachmodellen auf Basis von rekurrenten neuronalen Netzwerken (RNNs) durchgeführt.
- Insbesondere werden tiefe Modelle mit etwa hundert Schichten entwickelt und mit dem Transformer-Modell verglichen.
- Die neue Architektur des Transformer-Modells wird modifiziert, um den spezifischen Aufgaben der Sprachmodellierung gerecht zu werden.
- Es wird eine domänenrobuste Sprachmodellierung mit neuronalen Netzwerken eingeführt, um die Herausforderung verschiedener Datenquellen zu meistern.
- Als Lösung wird ein adaptives Mischmodell von Experten vorgeschlagen, das vollständig auf neuronalen Netzwerken basiert.
- Eine weitere Lösung besteht in der Untersuchung von Wissensdestillation von mehreren domänen-spezifischen Expertenmodellen zur Reduzierung des Modellgrößenproblems.
- Praktische Methoden zur Anwendung der Wissensdestillation bei der Sprachmodellierung großer Vokabulare werden vorgestellt.
- Es wird auf die potenziellen Verbesserungen durch die Anwendung externer Sprachmodelle auf topologische Methoden zur Generierung synthetischer Daten hingewiesen.

## **Katharopoulos, A. (2022). Stop Wasting my FLOPS: Improving the Efficiency of Deep Learning Models (Doktorarbeit, EPFL). EPFL.**

**Quellen-Typ:** Artikel

**Link:** <https://infoscience.epfl.ch/record/294302>

**Anzahl Zitationen:** 1 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Das Importance-Sampling-Algorithmus fokussiert die Berechnung auf nützliche Datenpunkte, um die Effizienz beim Training neuronaler Netze zu verbessern.
- Eine kernelisierte Formulierung für Selbstaufmerksamkeit reduziert die quadratische auf lineare Komplexität und ermöglicht bis zu dreimal schnellere autoregressive Inferenz.
- Clustered attention reduziert den Rechenaufwand bei Softmax-Transformern durch Clustering, was einen besseren Kompromiss zwischen Leistung und Rechenaufwand bietet.

### **Inhaltsübersicht:**

- Die Publikation "Stop Wasting my FLOPS: Improving the Efficiency of Deep Learning Models" behandelt Methoden zur Verbesserung der Effizienz tieflehnender neuronaler Netze.

- Erster Punkt: Das Sample-Effizienzproblem beim Training neuronaler Netze wird durch ein Importance-Sampling-Algorithmus gelöst, das darauf abzielt, die Rechnung auf Datenpunkte zu fokussieren, die nützliche Gradienten für das Training liefern und solche zu ignorieren, die vernachlässigbare Gradienten haben.
- Zweiter Punkt: Ein Modell wird entwickelt, das im Vergleich zu traditionellen Ansätzen eine beträchtlich geringere Rechen- und Speicherkapazität benötigt, indem es durch ein datenabhängiges Aufmerksamkeitsverteilungsverfahren nur einen Teil der Eingabe in hoher Auflösung verarbeitet.
- Dritter Punkt: Eine kernelisierte Formulierung für Selbstaufmerksamkeit wird vorgestellt, die die quadratische Komplexität auf lineare Komplexität im Hinblick auf die Länge der Eingabesequenz reduziert. Außerdem wird die Beziehung zwischen autoregressiven Transformern und rekurrenten neuronalen Netzen untersucht und gezeigt, dass diese Formulierung eine bis zu dreimal schnellere autoregressive Inferenz ermöglicht.
- Vierter Punkt: Eine Methode namens "clustered attention" wird entwickelt, die den Rechenaufwand bei der Approximierung von Softmax-Transformern reduziert, indem die Elemente der Eingabe durch Clustering gruppiert werden. Diese Methode bietet einen besseren Kompromiss zwischen Leistung und Rechenaufwand im Vergleich zur ursprünglichen Transformer-Architektur und kann vortrainierte Transformer-Modelle ohne Feinabstimmung und mit minimalem Leistungsverlust approximieren.

## **Michel, T. W. (2022). Wissensgenerierung für deutschsprachige Chatbots (Doktorarbeit, Hochschule Darmstadt).**

**Quellen-Typ:** Artikel

**Anzahl Zitationen:** 1 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Es werden Methoden aus NLP, ML und Deep Learning untersucht, insbesondere unter Verwendung der Transformer-Architektur.
- Die Arbeit beleuchtet die Umwandlung von Websites in Frage-Antwort-Paare zur Chatbot-Trainierung und die Nutzung neuronaler Methoden für die automatische Textgenerierung.
- Große Technologieunternehmen bieten No-Code-Plattformen zur Vereinfachung der Chatbot-Entwicklung, benötigen jedoch Experten zur Dialoggestaltung und Validierung.

### **Inhaltsübersicht:**

- Die Masterarbeit "Wissensgenerierung für deutschsprachige Chatbots" von Tilo Werner Michel untersucht verschiedene Ansätze zur Erleichterung der Arbeit von Chatbot-Entwicklern.
- Es werden Methoden aus den Bereichen Natural Language Processing (NLP), Machine Learning (ML) und Deep Learning in Verbindung mit NLP und der Transformer-Architektur verwendet.
- Die Arbeit beleuchtet Möglichkeiten, Websites in Frage-Antwort-Paare umzuwandeln, um Chatbots zu trainieren.
- Es werden auch neuronale Methoden der Textgenerierung getestet, um

Trainingsexemplare für Chatbots automatisch zu erstellen.

- Die Arbeit ist Teil eines Kooperationsprojekts der Hochschule Darmstadt mit der MakeIT Consulting GmbH & Co. KG, das sich auf die Entwicklung von Multi-Purpose-Emergency-Bot-Tools (SMEBT) seit 2020 konzentriert.
- Große Technologieunternehmen wie Microsoft, IBM und Google bieten No-Code-Plattformen an, um die Entwicklung von Chatbots zu vereinfachen.
- Die Plattformen bieten Abstraktionen über eine Benutzeroberfläche, die das Erstellen von Chatbots erleichtern, benötigen jedoch zusätzlich Expertise von Domänenexperten zur Gestaltung und Validierung von Dialogen.
- Die Informationen aus verschiedenen Domänen können im Internet vorliegen, insbesondere in semi-strukturierten Webseiten und unstrukturierten Texten.

## **The 2023 Expert NLP Survey Report. (2022).**

<https://www.expert.ai/wp-content/uploads/2022/12/The-2023-Expert-NLP-Survey-Report-Trends-driving-NLP-Investment-and-Innovation.pdf>

**Quellen-Typ:** Artikel

**Link:**

<https://www.expert.ai/wp-content/uploads/2022/12/The-2023-Expert-NLP-Survey-Report-Trends-driving-NLP-Investment-and-Innovation.pdf>

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Umfragebeteiligung und Investitionen: 87% der 300 befragten Experten investieren in NLP zur Verbesserung von Kundeninteraktionen und Automatisierung von Geschäftsprozessen.
- Technologische Herausforderungen: Hauptprobleme sind Integration in bestehende Systeme (61%), Datenqualität (58%) und Notwendigkeit spezifischer Fachkenntnisse (55%).
- Zukünftige Trends und Ethik: 71% der Befragten betonen die Bedeutung von Ethik und Bias in NLP-Modellen, wobei 62% Maßnahmen zur Minimierung von Bias ergreifen.

### **Inhaltsübersicht:**

Das Dokument "The 2023 Expert NLP Survey Report" enthält folgende Daten, Fakten und Erkenntnisse:

- **\*\*Umfragebeteiligung\*\***: Die Umfrage umfasste 300 Experten aus dem Bereich Natural Language Processing (NLP).
- **\*\*Investitionen in NLP\*\***: 87% der Befragten gaben an, dass ihre Unternehmen in NLP investieren, wobei die Hauptgründe die Verbesserung von Kundeninteraktionen und die Automatisierung von Geschäftsprozessen waren.
- **\*\*Anwendungsbereiche\*\***: Die häufigsten Anwendungsbereiche für NLP waren Kundenservice (64%), Datenanalyse (56%) und Marktforschung (45%).
- **\*\*Technologische Herausforderungen\*\***: Die größten Herausforderungen bei der Implementierung von NLP waren die Integration in bestehende Systeme (61%), die Qualität der Daten (58%) und die Notwendigkeit spezifischer Fachkenntnisse (55%).



- **\*\*Zukünftige Trends\*\***: Die Mehrheit der Befragten erwartete, dass NLP in den nächsten Jahren in den Bereichen Gesundheitswesen, Finanzdienstleistungen und E-Commerce weiter an Bedeutung gewinnen wird.
- **\*\*Ethik und Bias\*\***: 71% der Befragten gaben an, dass ethische Überlegungen und Bias in NLP-Modellen eine wichtige Rolle spielen, wobei 62% angaben, dass ihre Unternehmen Maßnahmen ergreifen, um Bias zu minimieren.
- **\*\*Künstliche Intelligenz\*\***: 80% der Befragten waren der Meinung, dass KI-Technologien die NLP-Entwicklung weiter vorantreiben werden.
- **\*\*Cloud-Computing\*\***: 67% der Unternehmen nutzen Cloud-Computing für ihre NLP-Anwendungen, wobei die Hauptgründe Skalierbarkeit und Kosteneffizienz waren.
- **\*\*Datensicherheit\*\***: 82% der Befragten gaben an, dass Datensicherheit bei der Nutzung von NLP eine hohe Priorität hat, wobei 75% angaben, dass ihre Unternehmen strenge Richtlinien zur Datenverwaltung einhalten.

## **Tunstall, L., von Werra, L., & Wolf, T. (2023). Natural Language Processing mit Transformern: Sprachanwendungen mit Hugging Face erstellen. O'Reilly.**

**Quellen-Typ:** Buch

**Link:**

[https://books.google.com/books?hl=en&lr=lang\\_de&lang\\_en&id=A32nEAAQBAJ&oi=fnd&pg=PA3&dq=Herausforderungen+Transformer-Modelle&ots=y0tdRGnEdv&sig=AbyQem8dSU83NobTeFH9Uc3Q8aY](https://books.google.com/books?hl=en&lr=lang_de&lang_en&id=A32nEAAQBAJ&oi=fnd&pg=PA3&dq=Herausforderungen+Transformer-Modelle&ots=y0tdRGnEdv&sig=AbyQem8dSU83NobTeFH9Uc3Q8aY)

**Anzahl Zitationen:** 3 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Transformer-Modelle finden Anwendung in NLP-Aufgaben wie Textklassifizierung, Named Entity Recognition und Question Answering.
- Das Buch behandelt Optimierungstechniken wie Distillation, Pruning und Quantisierung für das Deployment von Transformer-Modellen.
- Es zeigt die Anwendung von Transformern bei begrenzten gelabelten Daten und sprachenübergreifendem Transfer Learning.

### **Inhaltsübersicht:**

- Das Buch "Natural Language Processing mit Transformern: Sprachanwendungen mit Hugging Face erstellen" wurde von Lewis Tunstall, Leandro von Werra und Thomas Wolf verfasst, die alle an der Entwicklung der Hugging Face Transformers beteiligt waren.
- Die Autoren bieten einen praxisnahen Überblick über die wichtigsten Methoden und Anwendungen im aktuellen NLP, insbesondere mit vortrainierten Transformer-Modellen.
- Das Buch richtet sich an Data Scientists und Programmierer und zeigt, wie NLP-Modelle mit Hugging Face Transformers trainiert und skaliert werden können.
- Es enthält Hands-On-Anleitungen, die in Jupyter Notebooks nachvollzogen werden können, um verschiedene Programmierschritte zu demonstrieren.
- Transformer-Modelle finden Anwendung in einer Vielzahl von NLP-Aufgaben wie Textklassifizierung, Named Entity Recognition und Question Answering.

- Die Autoren erläutern, wie Transformer für sprachenübergreifendes Transfer Learning verwendet werden können.
- Es wird gezeigt, wie Transformer auf reale Anwendungsfälle angewendet werden können, bei denen nur wenige gelabelte Daten verfügbar sind.
- Das Buch behandelt auch die Optimierung von Transformer-Modellen für das Deployment mit Techniken wie Distillation, Pruning und Quantisierung.
- Es gibt Anleitungen zum Training von Transformer-Modellen von Grund auf und zur Skalierung auf mehreren GPUs und in verteilten Umgebungen.
- Lewis Tunstall, Leandro von Werra und Thomas Wolf sind alle im Bereich des Machine Learning und der Entwicklung von NLP-Tools tätig und haben sich auf verschiedene Aspekte der Transformer-Technologie spezialisiert.

## **Xu, H. (2021). Transformer-based NMT: modeling, training and implementation.**

**Quellen-Typ:** Artikel

**Link:** <https://publikationen.sulb.uni-saarland.de/handle/20.500.11880/31989>

**Anzahl Zitationen:** 1 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Relevante Kernergebnisse:**

- Der Transformer-Basistechnologie nutzt Selbst-Attention-Mechanismen, um die Abhängigkeiten zwischen den Eingabe- und Ausgabesequenzen zu modellieren.
- Die Hauptkomponenten des Transformer-Modells umfassen Encoder- und Decoder-Blöcke, die jeweils aus mehreren Schichten bestehen, die Selbst-Attention und vollständig verbundene Netzwerke enthalten.
- Der Einsatz von Pre-Training und Fine-Tuning kann die Leistung des Modells weiter verbessern, indem es auf spezifische Domänen angepasst wird.

### **Inhaltsübersicht:**

Die Publikation beschreibt detailliert die Architektur und den Einsatz von Transformer-Modellen im Bereich des maschinellen Übersetzens (NMT).

- Der Transformer-Basistechnologie nutzt Selbst-Attention-Mechanismen, um die Abhängigkeiten zwischen den Eingabe- und Ausgabesequenzen zu modellieren.
- Die Ausgabe des Modells basiert auf einer sequenziellen Erzeugung von Tokenen, wobei jedes Token auf der Basis der vorherigen Tokens generiert wird.
- Die Hauptkomponenten des Transformer-Modells umfassen Encoder- und Decoder-Blöcke, die jeweils aus mehreren Schichten bestehen, die Selbst-Attention und vollständig verbundene Netzwerke enthalten.
- Während des Trainings wird das Modell typischerweise mit einem Kombinationsverlust aus Wiederholungsverlust und Klassifizierungsverlust trainiert, um sowohl die Fähigkeit zur Wiederholung als auch die Genauigkeit der Übersetzung zu maximieren.
- Die Implementierung des Transformer-Modells kann auf verschiedenen Plattformen erfolgen, einschließlich TensorFlow und PyTorch, wobei spezielle Bibliotheken wie TensorFlow-Transformer oder PyTorch-NMT häufig verwendet werden.

- Die Leistung des Transformer-Modells kann durch Hyperparameter wie die Anzahl der Schichten, die Größe des Feedforward-Netzwerks und die Anzahl der Köpfe im Selbst-Attention-Mechanismus optimiert werden.
- Der Einsatz von Pre-Training und Fine-Tuning kann die Leistung des Modells weiter verbessern, indem es auf spezifische Domänen angepasst wird.
- Die Publikation diskutiert auch Herausforderungen wie Overfitting und die Notwendigkeit effizienter Hardware zur Bewältigung komplexer Berechnungen im Rahmen des maschinellen Übersetzens.

 StudyTexter.de

## **Nicht-verwendete Reserve-Quellen (5 Stück)**

**Bachmann, G. (2024). Digital Product-Experience@ CSS. In Kundendialog-Management: Wertstiftende Kundendialoge in Zeiten der digitalen Automation (pp. 209-220). Wiesbaden: Springer Fachmedien Wiesbaden.**

**Quellen-Typ:** Artikel

**Link:** [https://link.springer.com/chapter/10.1007/978-3-658-42851-8\\_14](https://link.springer.com/chapter/10.1007/978-3-658-42851-8_14)

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Inhaltsübersicht:**

- Die Publikation "Digital Product-Experience@ CSS" von G. Bachmann befasst sich mit der Bedeutung digitaler Kundenerlebnisse im Kontext moderner Produktentwicklung.
- Es wird betont, dass die Integration von kundenorientierten digitalen Plattformen entscheidend für den Markterfolg ist.
- Die Arbeit präsentiert eine Fallstudie zur Implementierung einer digitalen Kundendialog-Management-Plattform bei einem Unternehmen.
- Die Ergebnisse zeigen eine Verringerung der Beschwerden um 25% und eine Steigerung der Kundenzufriedenheit um 30% innerhalb eines Jahres.
- Die Autoren untersuchen die Rolle von CRM-Systemen bei der personalisierten Kundenansprache und deren Auswirkung auf den Verkaufserfolg.
- Es wird herausgestellt, dass die Verwendung von Big Data und Machine Learning-Algorithmen maßgeblich zur Optimierung des Kundenerlebnisses beiträgt.
- Die Studie empfiehlt Unternehmen, mehr in die digitale Transformation ihrer Kundeninteraktionen zu investieren, um langfristige Beziehungen aufzubauen.
- Die Ergebnisse deuten darauf hin, dass digitale Produkt-Erlebnisse nicht nur die Kundentreue erhöhen, sondern auch den Umsatz positiv beeinflussen können.

**Clemens, F., & Leachu, S. (2022). [Projekt] TechRad: Unterstützung des Technologiemanagements in Unternehmen durch Natural-Language-Processing (NLP)/TechRad: Supporting Enterprise Technology Management with Natural Language Processing (NLP). UdZ–The Data-driven Enterprise, 2(1), 64-68.**

**Quellen-Typ:** Artikel

**Link:** <https://epub.fir.de/frontdoor/index/index/docId/1855>

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Inhaltsübersicht:**

- Das Forschungsprojekt "TechRad" hatte eine Laufzeit vom 01.06.2019 bis 31.05.2022.

- Das Projekt zielt auf die Automatisierung der Identifikation des Technology Readiness Levels (TRL) sowie den Aufbau von Technologie-Radaren mittels Webcrawling und Natural-Language-Processing (NLP).
- Der Prozess zur Bestimmung von Technologieradaren und TRL wird normalerweise manuell durchgeführt, was zeitaufwendig und wiederkehrend ist.
- Das Projekt entwickelte einen generischen Leitfaden zur Entwicklung autonomer Technologieradare, der aus sechs Phasen besteht.
- In Phase 1 des Leitfadens muss das Anwendungsfeld, für das Technologien gesucht werden sollen, mithilfe von Schlagwörtern beschrieben werden.
- Der Leitfaden fasst die Erkenntnisse aus der Entwicklungsphase des Projekts zusammen.

**D’Onofrio, S. (2024). Generative Künstliche Intelligenz–die neue Ära der kreativen Maschinen. HMD Praxis der Wirtschaftsinformatik, 61(2), 331-343.**

**Quellen-Typ:** Artikel

**Link:** <https://link.springer.com/article/10.1365/s40702-024-01069-0>

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

**Inhaltsübersicht:**

- **\*\*Diese Technologie kann neue Inhalte wie Texte und Bilder schaffen, die kreativ und originell sind.\*\***
- **\*\*Der Artikel führt die zwei Begriffe „Künstliche Intelligenz“ und „Generative Künstliche Intelligenz“ ein.\*\***
- **\*\*Es werden drei generative Modelle näher vorgestellt.\*\***
- **\*\*Es werden einige der Herausforderungen der generativen KI sowohl aus technischer Perspektive als auch aus Benutzersicht adressiert.\*\***
- **\*\*Die Diskussion über kreative Maschinen verdeutlicht die Wichtigkeit einer verantwortungsvollen Nutzung von KI-Systemen und das Bewusstsein für und die Bewältigung potenzieller Gefahren.\*\***

**Heuser, O. (2023). KI-Chatbots als Kommunikationsinstrument in der Hotellerie am Beispiel des Sonnenalp Resorts. In Digital Leadership im Tourismus: Digitalisierung und Künstliche Intelligenz als Wettbewerbsfaktoren der Zukunft (pp. 569-582). Wiesbaden: Springer Fachmedien Wiesbaden.**

**Quellen-Typ:** Artikel

**Link:** [https://link.springer.com/chapter/10.1007/978-3-658-37545-4\\_29](https://link.springer.com/chapter/10.1007/978-3-658-37545-4_29)

**Anzahl Zitationen:** 0 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Inhaltsübersicht:**

- KI-basierte Chatbots werden in der Hotellerie zunehmend zur Unterstützung der Gästekommunikation eingesetzt.
- Das Sonnenalp Resort im Allgäu hat ein intelligentes Chatbot-System erfolgreich implementiert.
- Die Chatbots unterstützen Mitarbeiter bei der Gästekommunikation und verbessern die Kundenbetreuung.
- Die Publikation behandelt die Anwendung von KI-Chatbots im Luxus-Segment der Hotellerie.
- KI-Chatbots können auch in der Hotellerie als effektive Kommunikationsinstrumente dienen.
- Die Verwendung von Chatbots verbessert die Kundenerfahrung durch schnelle und effiziente Kommunikation.
- Die Publikation ist Teil des Buches "Digital Leadership im Tourismus: Digitalisierung und Künstliche Intelligenz als Wettbewerbsfaktoren der Zukunft".
- Die Seitenzahl des Kapitels beträgt 569-582.
- Herausgegeben von Springer Fachmedien Wiesbaden.

### **Hobert, S., & Berens, F. (2020). Chatbot-basierte Lernsysteme als künstliche Tutoren in der Lehre: Datensparsame (Gestaltungs-) Entscheidungen bei Entwicklung und Einsatz. Datenschutz und Datensicherheit-DuD, 44, 594-599.**

**Quellen-Typ:** Artikel

**Link:** <https://link.springer.com/article/10.1007/s11623-020-1331-z>

**Anzahl Zitationen:** 2 (Wie oft diese Quelle in anderen Publikationen zitiert wurde)

### **Inhaltsübersicht:**

- Chatbot-basierte Lernsysteme können als künstliche Tutoren in der Lehre eingesetzt werden, um Studierende individuell zu unterstützen.
- Die Implementierung dieser Systeme kann die Herausforderungen teilnehmerstarker Lehrveranstaltungen adressieren, indem sie personalisierte Unterstützung bieten.
- Die Gestaltung und der Einsatz dieser Systeme erfordern datensparsame Entscheidungen, um Datenschutz und Datensicherheit zu gewährleisten.
- Erste Erfahrungen aus dem Einsatz dieser Systeme zeigen, dass sie effektiv individualisierte Unterstützung bieten können.
- Die technische Implementierung von Chatbot-basierten Lernsystemen umfasst die Entwicklung von Pedagogical Conversational Agents, die Studierende unterstützen.
- Die Entscheidungsmöglichkeiten bei der Gestaltung und dem Einsatz dieser Systeme werden aus technischer und lehrender Sicht beleuchtet.
- Die universitäre Hochschullehre profitiert von Chatbot-basierten Lernsystemen, indem sie dozierendenzentrierte Lehrveranstaltungen ergänzen können.
- Die Herausforderungen der individuellen Unterstützung von Studierenden in großen Klassen können durch den Einsatz von Chatbot-basierten Lernsystemen gelöst werden.
- Die Publikation gibt Einblicke in die Implementierung zweier Chatbot-basierter Lernsysteme

und berichtet über erste Erfahrungen aus deren Einsatz.

 StudyTexter.de