



Cloud-basierte KI-Anwendungen für Natural Language Processing: Marktüberblick und Vorstellung von typischen Anwendungsfällen für den Mittelstand

Thorsten Pröhl  · Sabrina Baumgartner · Rüdiger Zarnekow

Eingegangen: 15. Juni 2023 / Angenommen: 10. Juli 2023 / Online publiziert: 22. August 2023
© The Author(s) 2023

Zusammenfassung Die Digitalisierung von Unternehmen schreitet voran und Künstliche Intelligenz (KI) ist längst nicht mehr nur ein Megatrend von morgen, sondern bereits im vielfältigen Einsatz. Große Unternehmen setzen bereits Lösungen ein, die auf Künstlicher Intelligenz basieren, um ihre Geschäftsprozesse sowie Abläufe zu automatisieren und zu optimieren. Um dem zunehmenden Wettbewerbsdruck standhalten zu können, müssen deshalb auch mittelständische Unternehmen mehr und mehr auf KI-Lösungen setzen. Ein Anwendungsfeld von KI stellt das Natural Language Processing dar, welches vielversprechende Möglichkeiten zur Automatisierung und Optimierung von Geschäftsprozessen von mittelständischen Unternehmen bereithält. KI-Anwendungen aus der Cloud bieten dabei den Vorteil, dass keine eigenen KI-Modelle entwickelt werden müssen, sondern auf bereits trainierte Modelle zurückgegriffen werden kann. Das erlaubt auch mittelständischen Unternehmen, die keine eigene Machine Learning Abteilung besitzen, KI in ihren Geschäftsprozessen einzusetzen. Im vorliegenden Beitrag werden typische Anwendungsfälle für den Mittelstand beschrieben und Anbieter cloud-basierter KI-Anwendungen für NLP übersichtsartig in Form eines Marktüberblicks präsentiert. Um den Mittelstand bei der Einführung von KI-Anwendungen zu unterstützen, werden typische Herausforderungen skizziert und ein Vorgehen zur Implementierung vorgeschlagen.

✉ Thorsten Pröhl · Sabrina Baumgartner
Trustami GmbH, Bamberger Straße 40, 10779 Berlin, Deutschland
E-Mail: tp@trustami.com

Thorsten Pröhl · Rüdiger Zarnekow
Fachgebiet Informations- und Kommunikationsmanagement, Technische Universität Berlin, Straße
des 17. Juni 135, 10623 Berlin, Deutschland

Schlüsselwörter Künstliche Intelligenz (KI) · Natural Language Processing (NLP) · Cloud-basierte KI · Anwendungsfälle · Herausforderungen · Vorgehensmodell

Cloud-based AI Applications for Natural Language Processing: Market Overview and Presentation of Typical Use Cases for Medium-sized Companies

Abstract The digitization of companies is advancing and artificial intelligence (AI) is no longer just a megatrend of tomorrow, but is already in widespread use. Large companies are already using solutions based on artificial intelligence to automate and optimize their business processes and workflows. In order to withstand the increasing competitive pressure, medium-sized companies must therefore also rely more and more on AI solutions. One field of application for AI is natural language processing, which offers promising opportunities for automating and optimizing the business processes of medium-sized companies. AI applications from the cloud offer the advantage that it is not necessary to develop their own AI models, but instead they can fall back on models that have already been trained. This also allows medium-sized companies that do not have their own machine learning department to use AI in their business processes. In this article, typical use cases for SMEs are described and providers of cloud-based AI applications for NLP are presented in the form of an overview of the market. In order to support SMEs in the introduction of AI applications, typical challenges are outlined and a procedure for implementation is proposed.

Keywords Artificial Intelligence (AI) · Natural Language Processing (NLP) · Cloud-based AI · Use Cases · Challenges · Process Model

1 Möglichkeiten von cloud-basierten KI-Anwendungen

Durch Cloud Computing können Unternehmen auf skalierbare und flexible IT-Infrastrukturen zugreifen. Es können Daten gespeichert, Server gemietet, Anwendungen gehostet, komplexe Services oder auch umfangreiche Geschäftsprozesse aus der Cloud konsumiert werden. Mit Hilfe von Cloud Computing realisieren Unternehmen Kosteneinsparungen (Hardware, Lizenzen, technisches Personal) und erhöhen ihre Flexibilität (Reagieren auf erhöhte Nachfrage oder kurzfristige Bedarfe) (Rittinghouse und Ransome 2009; Armbrust et al. 2010).

Künstliche Intelligenz (KI) wird inzwischen in allen Branchen, Sektoren und Bereichen in unterschiedlichem Umfang eingesetzt. Im Gesundheitswesen, der Automobilbranche oder der Finanzindustrie werden Fragestellungen mittlerweile datengetrieben und mit KI unterstützt beantwortet.

Deep Learning (DL) ist ein Teilbereich von Machine Learning (ML), wobei ML als Teilbereich von KI zu betrachten ist. Computer erhalten die Fähigkeit aus Erfahrungen zu lernen, Entscheidungen ohne konkrete Regeln zu treffen oder menschen-

ähnliche Tätigkeiten, wie z. B. Sprachinterpretation oder Bildanalyse auszuführen (Jordan und Mitchell 2015; Goodfellow et al. 2016).

Bestehende Anwendungen, Systeme und Prozesse können durch die Integration von KI erweitert werden, um den Grad der Automatisierung zu steigern, erhebliche Effizienzverbesserungen zu realisieren und die Robustheit von Prozessen zu erhöhen, indem auf unvorhergesehene (nicht den Regeln entsprechende) Fälle reagiert werden kann.

Die Kombination von KI-Anwendungen und Cloud Computing, also die Bereitstellung der KI-Anwendungen als cloud-basierte Lösungen, bieten Unternehmen jeder Größe den Zugriff auf leistungsstarke, elastische Ressourcen, um ihre Geschäftsprozesse und -anwendungen mit KI anzureichern, ohne selbst die notwendigen Ressourcen vorzuhalten, die es für das Training und die Ausführung von KI-Modellen bedarf. Unternehmen können direkt auf fertige KI-Anwendungen zugreifen ohne selbst die Schritte Datenbeschaffung, Annotation, Modellauswahl, Modelltraining, Modellevaluation und Modellinferenz auszuführen (LeCun et al. 2015).

Die bekannten Vorteile von Cloud Computing, wie Skalierbarkeit, erlauben es die KI-Anwendungen bedarfsorientiert einzusetzen, z. B. für die kurzfristige Textklassifikation oder Textanalyse, ohne eigene Ressourcen erst beschaffen, bereitzustellen und betreiben zu müssen.

Analog zu den Arten des Cloud Computing (u. a. SaaS, PaaS, IaaS) lassen sich cloud-basierte KI-Anwendungen unterscheiden. Es können fertige KI-Modelle (Sprachanalyse, Stimmungsanalyse, Bildererkennung, Übersetzung), KI-Datenanalyseplattformen oder Rechenressourcen für das eigene Modelltraining in Form von GPUs (Graphic Processing Unit) sowie TPUs (Tensor Processing Unit) aus der Cloud provisioniert werden. Der vorliegende Beitrag konzentriert sich auf fertige KI-Modelle, welche im Unternehmenskontext Geschäftsprozesse unterstützen. Ein weiterer Vorteil von der Nutzung fertiger, cloud-basierter KI-Anwendungen ist, dass die Modelle durch den Anbieter weiterentwickelt (trainiert) werden und so die Unternehmen von stetigen Verbesserungen (Prognosegenauigkeit, Robustheit und inhaltlicher Umfang) partizipieren.

Moderne KI-Anwendungen erlauben vielfältige Einsatzmöglichkeiten (Chen et al. 2012; Berente et al. 2021):

- Automatisierung wiederkehrender Tätigkeiten, wie die automatische Zuordnung von Daten, Datensätzen oder Buchungssätzen sowie die dynamische Überwachung von Schwellwerten, welche sonst manuell adjustiert werden müssen
- KI-gestützte Datenauswertung, indem große Mengen von Daten untersucht und nicht vorher definierte Muster erkannt werden, die wiederum Rückschlüsse auf Trends und Wirkungszusammenhänge zulassen
- Sprachverarbeitung, also die Nutzung von Natural Language Processing (NLP), um Texte zu analysieren, zu verstehen und zu verarbeiten, u. a. mit dem Ziel Textzusammenfassungen zu erstellen oder Dialogsysteme zu entwickeln
- Analyse von Bild- und Videodaten, um etwa Mediziner bei der Bildanalyse (Röntgen, MRT) zu unterstützen oder Personen automatisch in Videostreams zu identifizieren

- Vorhersagemodelle, die sowohl historische als auch aktuelle Daten verarbeiten, um Kundenverhalten, Fertigungsprozesse oder Verkehrsflüsse zu prognostizieren
- Personalisierte Empfehlungen, welche Nutzern, Kunden oder Unternehmen aufgrund von historischen Präferenzen oder Auswahlen Einkaufsvorschläge im E-Commerce unterbreiten
- Betrugserkennung zur Identifikation von verdächtigen Aktivitäten und handelnden Personen im Kontext von Bewertungen, Online-Einkäufen oder Kreditkarten Transaktionen

Cloud-basierte KI-Anwendungen bieten vielfältige Möglichkeiten, um Unternehmen in ihren Geschäftsprozessen zu unterstützen, die Wettbewerbsfähigkeit zu steigern oder Erkenntnisse aus Daten zu gewinnen.

2 Grundlagen von Natural Language Processing

NLP stellt einen Bereich von KI an der Schnittstelle zur Linguistik dar. Mit Hilfe von NLP wird Sprache durch Computer analysiert und verarbeitet. Dabei werden die NLP-Techniken und -Konzepte mit dem Ziel eingesetzt, Computern die Fähigkeit zu geben, menschliche Sprache zu analysieren, zu interpretieren und zu generieren (Manning und Schütze 2005; Jurafsky und Martin 2009; Goldberg 2022). Hierbei spielen Maschinelles Lernen und Deep Learning sowie moderne Sprachmodelle eine entscheidende Rolle. Gerade bei der Textgenerierung werden große Sprachmodelle (engl. Large Language Models), wie RoBERTa, ChatGPT oder GPT4, eingesetzt.

Einsatzgebiete von NLP sind vielfältig und umfassen Übersetzungen, Klassifikationen von Texten, Stimmungsanalysen, Chatbots oder aber auch Text-zu-Sprache-Systeme. Im Kontext von NLP wird Sprache auf der Ebene von Sätzen oder einzelnen Wörtern verarbeitet. Dabei stellen die folgenden Techniken grundlegende Funktionalitäten dar. Bei der Tokenization wird ein Satz in einzelne Wörter oder Wortkombinationen aufgeteilt (Grefenstette 1999). Daneben stellt das Part-of-Speech-Tagging (POS) die Analyse und Zuordnung von Wortarten, wie z.B. Substantiven und Verben, dar. Darüber hinaus ist es möglich mit Named Entity Recognition (NER) Entitäten in Sätzen zu finden. Entitäten können z.B. Personen, Orte, Marken oder Organisationen sein. Weiterhin stellt das sogenannte Parsing, also die Analyse der Satzstruktur, eine grundlegende Technik dar.

Eine weitere komplexere Technik des NLPs ist zum Beispiel die Stimmungsanalyse (engl. Sentiment Analysis). Hier werden Stimmungen und Gefühle von Texten erkannt. Oftmals handelt es sich um eine Polaritätsanalyse, wobei die Polarität von negativ, über neutral, hinzu positiv ausfallen kann. Bei einer tatsächlichen Sentiment Analyse steht neben der Polarität auch der Grad der Subjektivität im Fokus der Analyse. Jedoch werden die Begriffe Polaritäts- und Sentiment Analyse häufig identisch verwendet. Mit Hilfe der Stimmungsanalyse werden beispielsweise Meinungen von Kunden in E-Commerce Bewertungen automatisch ausgewertet (Manning und Schütze 2005; Dashtipour et al. 2016).

Im Rahmen des Topic Modeling werden Themen in Texten identifiziert und erkannte Schlüsselwörter zu den identifizierten und passenden Themen zugeordnet.

Typische Verfahren für das Topic Modeling sind Latent Dirichlet Allocation (LDA) oder transformer-basierte Ansätze, die z. B. auf BERT-Sprachmodelle zurückgreifen. Mit Topic Modeling können Trends in Social Media oder Nachrichtenartikeln analysiert werden (Jelodar et al. 2019).

Häufig müssen lange Texte zusammengefasst und der Inhalt auf wenige Sätze kondensiert werden. Durch Text Summarization können Zusammenfassungen von Texten erstellt werden. Dabei kann zwischen Extractive und Abstractive Text Summarization unterschieden werden. Die Extractive Summarization wählt einzelne relevante Sätze aus, ohne diese zu verändern und kreiert darüber einen neuen Text, wohingegen die Abstractive Summarization neuen Text generiert. Ein Anwendungsgebiet stellt die Zusammenfassung von Nachrichten zu einem Thema dar (Allahyari et al. 2017).

In Sprachassistenten wie Siri oder Alexa werden Techniken und Algorithmen verwendet, um die gesprochene Sprache des Menschen zu verstehen und entsprechende Antworten geben zu können. Im umgekehrten Fall wird geschriebener Text, z. B. Hörbücher oder vorhandene Informationen, in natürliche Sprache umgewandelt (Text-to-Speech). Anwendung findet dies nicht nur in Sprachassistenten, sondern auch bei Hörbüchern oder Vorlesefunktionen auf Webseiten (Jurafsky und Martin 2009).

Um semantische Ähnlichkeiten von Wörtern, Sätzen oder Texten zu bestimmen, werden Word Embeddings verwendet, bei denen Wörter in Form von Vektoren ausgedrückt werden. Semantische Ähnlichkeit bietet gegenüber phonetischer Ähnlichkeit den Vorteil, nicht auf Wortstämme oder Teilstrings (Longest Common Subsequence) begrenzt zu sein (Lebret 2016).

Neben dem Parsing von Sätzen wird das Dependency Parsing verwendet, um syntaktische Beziehungen von Wörtern in einem Satz zu analysieren. Daneben stellt die Coreference Resolution ein Verfahren dar, um den Bezug von Wörtern in Texten zu bestimmen. Dabei wird untersucht, welche Wörter sich auf dieselbe Person oder dasselbe Objekt beziehen (Rahman und Ng 2009).

Ein großes Thema stellt die Sprachgenerierung dar. Es werden Sprachmodelle, wie GPT-2 oder FastText, eingesetzt, um automatisch Texte zu erstellen, die wirken, als wenn sie von Menschen verfasst worden wären. Gerade das Aufkommen von großen Sprachmodellen, wie GPT-3, BERT, T5 oder RoBERTa, lassen die Sprachgenerierung noch erheblich natürlicher erscheinen (Wei et al. 2023). Aktuelle LLMs wie ChatGPT oder GPT-4, greifen auf einen wesentlich größeren Trainingsdatenbestand zurück, sodass auch ein besseres Verstehen von Nutzeranfragen gewährleistet wird. Gerade GPT-4 ist multimodal und lernt damit nicht nur aus Texten, sondern auch aus Bildern.

Die vorliegenden ausgewählten Techniken geben einen Eindruck von den Möglichkeiten und Aufgaben, die mit NLP bearbeitet werden.

3 Typische Anwendungsfälle für den Mittelstand

In mittelständischen Unternehmen können NLP-Techniken, -Konzepte und -Verfahren für unterschiedliche Anwendungsfälle eingesetzt werden. NLP bietet sich im

Kontext von mittelständischen Unternehmen an, da hier aufgrund von begrenzten personellen Ressourcen Automatisierungskonzepte zu einer Entlastung führen und effiziente Prozesse für das Bestehen im Wettbewerb notwendig sind. In Abgrenzung zu Großunternehmen verfügen mittelständische Unternehmen nicht unweigerlich über Data Science oder Machine Learning Personal, wodurch für diese Unternehmen der Einsatz von vortrainierten KI-Modellen oder existierenden KI-Anwendungen aus der Cloud ökonomischer ist und einen Startpunkt für eigene Machine Learning Aktivitäten darstellt.

Im Folgenden werden sieben typische NLP-Aufgaben im Kontext möglicher Anwendungsfälle in mittelständischen Unternehmen beschrieben. Insbesondere werden die Möglichkeiten von Language Detection, Sentiment Analysis, Text Summarization, Topic Extraction, Keyword Analysis, Named Entity Recognition und Sentence Tokenization beleuchtet (Khurana et al. 2023).

Language Detection: Im Speziellen können Unternehmen auch im Support NLP einsetzen, um automatisch die Sprache von Texten, wie Kundenanfragen, E-Mails, Chats oder Tickets, zu bestimmen. Durch die Sprachbestimmung (engl. Language Detection) können die eingehenden Texte an den richtigen Ansprechpartner, der die Sprache beherrscht, weitergeleitet und durch diesen beantwortet werden. Das steigert die Effizienz und Kundenzufriedenheit (Rajanak et al. 2023).

Sentiment Analysis: Mittelständische Unternehmen können von Stimmungsanalysen (engl. Sentiment Analysis) vielfältig partizipieren. Gerade im Marketing können Stimmungen, Meinungen und Reaktionen von Kunden automatisiert analysiert und bewertet werden, sodass durch Nutzung des Kundenfeedbacks neue oder verbesserte Produkte wie auch Dienstleistungen entstehen können. Weiterhin dienen diese Analysen dazu, einen Wechsel in der Stimmung oder das Aufziehen von Krisen im Umfeld des Unternehmens frühzeitig zu erkennen. Nachrichten, Google Einträge oder Social-Media-Beiträge können das Image des Unternehmens negativ beeinflussen, sodass frühzeitige und automatische Analysen unabdingbar sind, um schnell auf diese reagieren zu können. Auch das Kundenerlebnis insgesamt profitiert von Stimmungsanalysen, da Feedback analysiert und auf den Kunden direkt eingegangen werden kann, was ohne automatische Analysen in einem solchen Umfang schlicht unmöglich ist. Des Weiteren können Priorisierungen im Support, z. B. bei E-Mails oder Tickets vorgenommen werden, indem die Stimmung der Texte zugrunde gelegt wird, sodass vorrangig die negativen Anfragen bearbeitet und positive Rückmeldungen zu einem späteren Zeitpunkt beachtet werden. Schließlich können Sentiment Analysen Stimmungen in Finanznachrichten, Berichten zu Unternehmen oder Finanzdokumenten erkennen und damit Trends identifizieren. Mit Hilfe dieser zusätzlichen Informationsquelle können Prognosemodelle erweitert und verbessert werden (Neri et al. 2012; Dashtipour et al. 2016).

Text Summarization: Lange Nachrichten- oder Artikeltexte können durch die Nutzung von Text Summarization zusammengefasst werden, sodass eine schnelle Übersicht über die wesentlichen Inhalte möglich wird und nicht seitenweise Text gelesen und erfasst werden muss. Bei der Verwaltung von E-Mails können Zusammenfassungen die Mitarbeitenden bei der Priorisierung der Abarbeitung von E-Mails unterstützen. Trotz langer E-Mails kann der Bearbeitende den Inhalt der E-Mail schnell erfassen und entscheiden, welche Wichtigkeit diese hat. Auch Anleitungen

und technische Dokumentationen können durch Text Summarization zusammengefasst werden. Damit bleiben die wesentlichen Informationen erhalten und die Nutzer können effizienter und schneller mit diesen Dokumenten umgehen sowie auf die wichtigen Informationen zurückgreifen (Allahyari et al. 2017).

Topic Extraction: NLP bietet durch Topic Extraction die Möglichkeit Schlüsselwörter in Texten zu finden und in Themen zu gruppieren. Diese Schlüsselwörter werden durch das Machine Learning selbstständig und automatisch geordnet und es bedarf keiner manuellen Zuweisung. Als Anwendungsfall für dieses Verfahren ist die Trendanalyse zu nennen, bei der Entwicklungen in einer großen Menge von Daten wie Beiträgen identifiziert werden. Beispiele hierfür sind Beiträge in Blogs, auf Nachrichtenseiten oder im Social-Media-Bereich. Neben der Trendanalyse ist auch die Inhaltsanalyse ein Anwendungsgebiet der Topic Extraction. Hier werden relevante Themen und Inhalte aus verschiedenartigen Texten (wie Berichte oder Bücher) extrahiert. Im Kundenservice können die einzelnen Anfragen automatisch Themenblöcken zugeordnet und dementsprechend passenden Mitarbeitenden zur Weiterverarbeitung übergeben werden. Des Weiteren kann Topic Extraction für die Kuratation von Inhalten, z. B. in Foren oder auf Seiten, die Online-Bewertungen enthalten, verwendet werden. Unerwünschte oder problematische Themen werden zunächst automatisch identifiziert und können dann einer manuellen Prüfung übergeben werden (Velden et al. 2017).

Keyword Analysis: Als Grundlage für die Topic Extraction werden Keyword Analysen (engl. Keyword Analysis) verwendet. Durch die Nutzung von Machine Learning werden relevante Schlüsselwörter in Texten identifiziert. Unternehmen können diese Art der Analyse z. B. für die Suchmaschinenoptimierung (SEO) einsetzen, um relevante Schlüsselwörter auf konkurrierenden Webseiten, Blogs und Online Shops automatisch zu erkennen. Diese Keywords werden dann gezielt auf den eigenen Seiten platziert, um das Suchmaschinenranking zu verbessern. Aber auch das Marketing von Unternehmen kann durch Keyword Analysen verbessert werden, indem wichtige Schlüsselwörter und Begriffe gefunden werden, die in Marketing-Kampagnen und auch Anzeigen genutzt werden, um passende Zielgruppen besser zu erreichen. Im Support liefern Keyword Analysen automatische Einblicke in die Inhalte von Anfragen. Es können Filterungen im Mailverkehr auf spezielle Keywords, wie z. B. „Beschwerde“, „Wichtig“, „Dringend“, „Ausfall“ oder „Vorgesetzter“, eingestellt werden, wodurch eine Priorisierung dieser Anfragen möglich ist (Onan et al. 2016).

Named Entity Recognition: Um mehr Verständnis und Erkenntnisse aus Texten zu bekommen, können Entitäten, also Personen, Orte oder Marken, bestimmt werden. Mit Hilfe der Named Entity Recognition (NER) werden diese Entitäten in Dokumenten identifiziert. Dabei stellt NER nicht eine einfache Wortliste dar, weil sonst beispielsweise „Mars“ nur ein Planet (Ort) und kein Schokoriegel (Marke) wäre. Durch Machine Learning Modelle werden die Entitäten in ihrem jeweiligen Kontext betrachtet. Bei der Extraktion von Informationen spielt NER eine wichtige Rolle. In unstrukturierten Texten, also z. B. Nachrichtenartikel oder Webseiten, werden mit NER Entitäten wie Autoren, Prominente oder Orte gefunden. Daneben hilft NER auch bei der Übersetzung von Texten, da gerade Namen oder Orte, wie auch Organisationsnamen in Übersetzungen gleichbleiben oder explizit ihrer

Übersetzung zugeführt werden müssen. Weiterhin wird NER in Chatbots, also bei Kundeninteraktionen verwendet, um automatisch Produkte oder Services des eigenen Unternehmens zu identifizieren und im Chat darauf Bezug zu nehmen, z. B. durch das Anzeigen von passenden FAQ und Hilfeseiten oder beim Finden eines Ansprechpartners (Mohamad Suhaili et al. 2021; Li et al. 2022).

Sentence Tokenization: Eine grundlegende, aber doch für viele NLP-Aufgaben notwendige Funktion stellt die Sentence Tokenization dar. Es geht um die Zerlegung eines Textes in einzelne Sätze und ist in Abgrenzung zur Tokenization, also der Zerlegung eines Satzes in einzelne Wörter oder Tokens zu verstehen. Im Kontext von Übersetzungen wird Sentence Tokenization benutzt, um Dokumente in Absätze oder Sätze zu zerlegen, wodurch die Übersetzung einfacher wird. Bei der Informationsextraktion findet dieses Verfahren auch Anwendung, da in einem Dokument zahlreiche Informationen vorhanden sind, die sich lediglich auf Sätze oder Absätze beziehen, und nicht auf das gesamte Dokument. Weiterhin werden in Textanalysen Satzauftrennungen verwendet, um z. B. die Anzahl der Sätze in einem Dokument oder die Länge der Sätze zu bestimmen, was häufige Metriken sind, die einen Autor oder Verfasser beschreiben (Grefenstette 1999).

Die dargestellten Szenarien geben einen Eindruck von den vielfältigen NLP-Einsatzmöglichkeiten im Kontext von mittelständischen Unternehmen.

4 Marktüberblick zu cloud-basierten KI-Anwendungen für NLP

Die skizzierten Anwendungsfälle und Einsatzmöglichkeiten der ausgewählten NLP-Techniken bieten mittelständischen Unternehmen weitreichende Potenziale. Um den großen Markt von KI-Anwendungen zu überblicken, wird ein Marktüberblick erstellt, der sich auf KI-Anwendungen für NLP konzentriert und sich besonders auf die ausgewählten Techniken fokussiert. Damit wird es mittelständischen Unternehmen möglich, die aufgezeigten Anwendungsfälle und Einsatzmöglichkeiten im eigenen Unternehmenskontext zu realisieren.

Für die Anfertigung des Marktüberblicks wurden im ersten Schritt relevante Schlagwörter und Suchbegriffe aus den Domänen KI und NLP identifiziert. Die Begriffe wurden sowohl auf Englisch als auch Deutsch mit der Suchmaschine Google gesucht. Neben den allgemeinen Suchbegriffen wie beispielsweise „Künstliche Intelligenz API“, „Natural Language Processing API“ oder „Machine Learning API“, wurden auch spezielle Suchbegriffe, die die einzelnen Dienste an sich beschreiben, genutzt. Dazu gehören etwa „Sentiment Analysis“, „Stimmungsanalyse“, „Polaritätsanalyse“ in Kombination mit Wörtern wie „API“, „Service“, „Tool“ oder „Online“. Im zweiten Schritt wurden die 41 identifizierten Anbieter cloud-basierter KI-Anwendungen für NLP näher untersucht. Nach der ersten Analyse wurden 32 Anbieter als relevant eingestuft. Es kommen 14 Anbieter aus den USA, 14 aus der EU sowie vier aus anderen Ländern. Abb. 1 stellt die näher untersuchten Unternehmen überblickartig dar. Es werden die relevanten Kriterien aufgezeigt. Neben dem Sitz des Unternehmens wird als Kriterium auch angegeben, ob das Unternehmen in Deutschland oder Europa tätig ist. Außerdem ist zu erkennen, ob Deutsch als Sprache für die NLP-Anwendungen zur Verfügung steht. Des Weiteren wurde untersucht,

| Unternehmen | Sitz | In D/Europa tätig | Deutsch unterstützt | Language Detection | German Sentiment Analysis | Text Summarization | Topic Extraction | Keywords Analysis | Named Entity Recognition | Sentence Tokenization |
|----------------------------|-------------|-------------------|---------------------|--------------------|---------------------------|--------------------|------------------|-------------------|--------------------------|-----------------------|
| Alteryx | USA | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | |
| Amazon Web Services | USA | ✓ | ✓ | ✓ | ✓ | | | | ✓ | |
| Aylii | Irland | ✓ | ✓ | | | | | ✓ | ✓ | |
| Caplena | Schweiz | ✓ | ✓ | | ✓ | | | | | |
| Clarifai | USA | | | | | | | | ✓ | |
| Deep.ai | USA | | | | ✓ | ✓ | | ✓ | | |
| Detect Language API | Litauen | ✓ | ✓ | ✓ | | | | | | |
| Expert.ai | Italien | ✓ | ✓ | | ✓ | ✓ | | | ✓ | |
| Google Cloud | USA | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | |
| IBM | USA | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | |
| Intellexer (EffectiveSoft) | USA | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | |
| Komprehend (ParallelDots) | USA | | ✓ | | ✓ | | | ✓ | ✓ | |
| LanguageLayer API | Österreich | ✓ | ✓ | ✓ | | | | | | |
| Lexalytics | USA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ |
| MeaningCloud | USA | | ✓ | ✓ | | ✓ | | | | |
| Microsoft | USA | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| MonkeyLearn | Argentinien | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | |
| Neural Space | England | ✓ | ✓ | ✓ | | | | | ✓ | |
| NLP Cloud | Frankreich | ✓ | ✓ | ✓ | ✓ | ✓ | | | | |
| Oneai | USA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| PangeaMT | Spanien | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | |
| Paraphraser.io | Pakistan | ✓ | ✓ | | | ✓ | | | | |
| Plnia | USA | | ✓ | ✓ | | ✓ | | ✓ | | |
| Repustate | Kanada | ✓ | ✓ | | ✓ | | | | ✓ | |
| SummarizeBot | Lettland | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |
| Symanto | Deutschland | ✓ | ✓ | | ✓ | | ✓ | | | |
| Text2data | Polen | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | |
| Textrazor | England | ✓ | ✓ | ✓ | | | | | ✓ | |
| Translated Labs | Italien | ✓ | ✓ | ✓ | | | | | | |
| Trustami AI | Deutschland | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Twinword | USA | | | | | | | ✓ | | |
| Yonder Labs | Italien | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | |

Abb. 1 Marktüberblick über 32 Anbieter cloud-basierter KI-Anwendungen für NLP

ob die unterschiedlichen Anbieter die sieben NLP-Services, deren Anwendungsszenarien zuvor beschrieben wurden, unterstützen. Gerade bei der Sentiment Analyse wurde das Kriterium härter ausgelegt, indem dort von „German Sentiment Analyse“ gesprochen wird. Es finden sich deutlich mehr Stimmungsanalysen, die primär Englisch als Sprache unterstützen, was sicherlich an dem geringeren Umfang von Trainingsdaten für die deutsche Sprache liegt.

Eine Vielzahl von Anbietern stammt aus Nicht-EU-Mitgliedsstaaten. Bei diesen Anbietern sollte besonderes Augenmerk auf die Einhaltung regulatorischer Anforderungen, wie insbesondere Datenschutz, gelegt werden. Dabei müssen auch Veränderungen hinsichtlich Datenschutzabkommen zwischen Staaten beachtet werden. Weiterhin sollte bei der Wahl des Anbieters berücksichtigt werden, ob die Input Daten von dem Anbieter für das Modelltraining verwendet werden. Hier könnten Probleme im Kontext von personenbezogenen Daten oder z. B. medizinischen, patientenorientierten Dokumentationen entstehen.

Aus der Übersicht lässt sich schnell entnehmen, dass besonders viele Anbieter die Services Sentiment Analysis, Named Entity Recognition und Language Detection offerieren. Interessanterweise wird die Sentence Tokenization nur von sehr wenigen Unternehmen angeboten, wenngleich diese einen notwendigen Baustein für NLP-Workflows darstellt.

Bei der Untersuchung der Anbieter sind ganz unterschiedliche Abrechnungsmodelle ersichtlich geworden. Neben klassischen request-basierten Abrechnungsmodellen, sind auch character-, token- oder traffic-basierte Preismodelle zu finden. Gerade bei den token-basierten Abrechnungsmodellen, die z. B. von Anbietern generativer KI-Modelle bekannt ist, werden sowohl die Input als auch die Output Tokens für die Berechnung zugrunde gelegt. Weiterhin kann man bei dem Vergleich der Unternehmen erkennen, dass die Anbieter ihre KI-Anwendungen in unterschiedlicher Weise bereitstellen. So gibt es Anbieter, die nur vorgefertigte Pakete hinter einer API kapseln und somit einfache NLP-Services anbieten, dem gegenüber gibt es auch Anbieter, die selbsttrainierte, Deep Learning basierte KI-Modelle für die Lösung der NLP-Aufgaben zur Verfügung stellen. Schließlich bieten einige Anbieter vorgefertigte Software Frameworks (auch Konnektoren genannt) für verschiedene Programmiersprachen an, um Eintrittsbarrieren bei der Nutzung der NLP-Services zu minimieren. Zudem unterscheiden sich die Klassifikationsgenauigkeiten der eingesetzten Machine Learning Modelle von Anbieter zu Anbieter und können vor allem beim Einsatz in unterschiedlichen fachlichen Domänen variieren. Bemerkenswert ist zudem, dass die Provider einen sehr unterschiedlichen Umfang an NLP-Services anbieten.

5 Herausforderungen und Vorgehen zur Anbieterswahl

Es ergeben sich bei der Auswahl von cloud-basierten KI-Anwendungen für NLP spezifische Herausforderungen, die mit Hinblick auf den angestrebten Projekterfolg berücksichtigt werden sollten (Khurana et al. 2023).

Bei der Nutzung cloud-basierter KI-Anwendungen findet oft ein Austausch sensibler Daten zwischen Anbieter und Kunde statt, sodass Datenschutz und Datensi-

cherheit gewährleistet sein müssen. Daneben spielt die Verarbeitung großer Datenmengen eine wichtige Rolle, wenn die NLP-Anwendung erst voll im Unternehmen integriert ist, sodass darauf geachtet werden muss, dass die erforderlichen Ressourcen beim Anbieter vorhanden sind und somit Skalierbarkeit gegeben ist. Nicht nur Skalierbarkeit, sondern auch Flexibilität muss gewährleistet sein. Abhängig vom spezifischen Anwendungsfall und der Domäne können die Anforderungen an die NLP-Anwendung stark unterschiedlich sein, sodass diese anpassbar oder personalisierbar sein muss. Gerade sprachliche und kulturelle Anpassungen sollten möglich sein, um verschiedenen Sprachen, Dialekten oder kulturellen Unterschieden gerecht zu werden. Daneben sollte eine Anpassung an verschiedene Domänen oder Branchen realisierbar sein, was mitunter eine Herausforderung darstellen kann. Sollte das Erfordernis von Echtzeitverarbeitung im Kontext des Anwendungsfalls vorhanden sein, z. B. für Spracherkennungssysteme oder Chatbots, so sollte der Anbieter diese Echtzeitfähigkeiten ermöglichen und unterstützen, wobei hier Antwortzeiten zu berücksichtigen sind. Zur Bestimmung der Kosten der cloud-basierten KI-Anwendung sollte im Vorhinein eine Analyse des Preismodells und einhergehender Kosten stattfinden. Eine Unwägbarkeit stellt der flexible Anteil der Kosten, also die nutzungsabhängigen Kosten, dar. Unternehmen können im Vorfeld ihr Nutzungskontingent meist nur schwer prognostizieren, weshalb zwangsweise mit Limits (Hard Limits) gearbeitet werden sollte. Darüber hinaus ist die Bestimmung der Modellqualität und Modellzuverlässigkeit ein wichtiger Erfolgsfaktor für die NLP-Anwendung. Die Quantifizierung stellt bei cloud-basierten KI-Anwendungen eine Herausforderung dar. Cloud-basierte KI-Anwendungen für NLP nutzen komplexe Deep Learning Modelle, die schwer zu interpretieren sind. Die Entscheidungsfindung und Funktionsweise dieser Modelle zu verstehen, kann herausfordernd sein, ist allerdings im Kontext von regulatorischen Anforderungen erforderlich.

Um die vorliegenden Herausforderungen zu adressieren, sollten die folgenden fünf Schritte Berücksichtigung bei der Wahl eines Anbieters und der Nutzung von KI-Anwendungen für NLP finden. Das vorliegende Modell eignet sich für die Auswahl eines Anbieters cloud-basierter NLP-Anwendungen, da die NLP-spezifischen Herausforderungen in den einzelnen Schritten adressiert werden (Abb. 2).

1. Analyse und Identifizierung von Anforderungen

Im ersten Schritt wird das eigene Unternehmen, insbesondere die Geschäfts- und unterstützenden Prozesse, im Hinblick auf Verbesserungs-, Automatisierungs- und Optimierungspotenziale analysiert. Darauf basierend werden dann konkrete Anforderungen für eine NLP-Unterstützung abgeleitet und identifiziert, wie NLP-Funktionen die Prozesse verbessern können. Bei der Bedarfsermittlung werden die spezifischen Anforderungen für den NLP-Einsatz identifiziert und aufgelistet. Dabei sollten insbesondere Sprachunterstützung, Art der Daten, Anwendungsdomäne, Mengenge-



Abb. 2 Vorgehen zur Auswahl eines Anbieters cloud-basierter KI-Anwendungen für NLP

rüste, Sicherheit und Interoperabilität betrachtet werden. Gerade der Domänenbezug der eigenen Daten ist für die spätere Auswahl des Anbieters wichtig, da NLP- oder Sprachmodelle z. B. Schwierigkeiten mit Fachwörtern aufweisen können, wie etwa bei medizinischen oder technischen Dokumentationen sowie Rechtstexten. Je nach Art der Daten sind außerdem unterschiedliche Sicherheitsstandards oder regulatorische Anforderungen zu berücksichtigen. Im Rahmen einer Kosten-Nutzen-Analyse sind die gesamten Kosten für die Nutzung von externen Diensten zu berücksichtigen, welche sich u. a. aus Integrations-, Anpassungs-, Lizenz- und Schulungskosten zusammensetzen. Diese gilt es in Relation zu den erwarteten Vorteilen und Nutzen (höherer Grad der Automatisierung, schnellerer Prozessablauf oder höhere Kundenzufriedenheit) zu setzen und diese zu quantifizieren, also einen Business Case aufzustellen.

2. Identifikation und Bewertung von NLP-Anbietern

Im nächsten Schritt werden cloud-basierte NLP-Anbieter recherchiert und identifiziert. Dabei spielen neben einer klassischen Online Suche auch Bewertungsportale und Vergleichslisten eine entscheidende Rolle. Bei der Wahl und Entscheidung hinsichtlich Eignung können auch Referenzen sowie Case Studies einen Einblick in die Möglichkeiten und bisherigen Implementierungen geben. Die Anbieter werden hinsichtlich der vorab identifizierten Anforderungen bewertet, um eine erste Vorauswahl (Shortlist) zu treffen. Neben den eigenen Anforderungen sollten auch allgemeine Aspekte wie Zuverlässigkeit, Verfügbarkeit, Servicequalität sowie Supportfähigkeit der Anbieter berücksichtigt werden. Weiterhin sind die folgenden speziellen Aspekte zu beachten: Tarifgestaltung inkl. Preise für Einmalaufwände, Datenschutz, Integration in vorhandene Technologieinfrastruktur, Verfügbarkeit und Dokumentation von APIs, Qualität der ML- und NLP-Modelle sowie Domänenspezialisierung. Um die Zukunftssicherheit des NLP-Anbieters zu bewerten, sollte vor allem geprüft werden, ob dieser Innovationstätigkeiten wie eigene Forschung oder Weiterentwicklung der Modelle unternimmt. Auch die Prüfung hinsichtlich Skalierbarkeit spielt eine wichtige Rolle. Denn der Anbieter soll die eigenen, tendenziell im Zeitablauf steigenden Bedarfe, abdecken können.

3. Durchführung eines Proof of Concept

Im dritten Schritt sollte ein Proof of Concept (PoC) im eigenen Unternehmen durchgeführt werden. Dabei wird die tatsächliche Eignung der NLP-Anwendung im Kontext der eigenen Geschäftsprozesse betrachtet und analysiert. Dieser Schritt dient der praktischen Evaluierung der NLP-Lösung und ist mit minimalen Kosten verbunden. Typischerweise werden ein oder zwei Anbieter aus der vorher aufgestellten Shortlist für den PoC verwendet. Man testet die NLP-Fähigkeiten und Funktionen im Kontext der eigenen Daten, Prozesse und Infrastruktur. Die Erkenntnisse und Ergebnisse des Proof of Concept sollten übersichtsartig dokumentiert werden.

4. Bewertung der Ergebnisse und Entscheidungsfindung

In diesem Schritt werden die dokumentierten Ergebnisse aus der Durchführung des PoC mit dem eingangs aufgestellten Business Case abgeglichen, um eine Entscheidung für einen oder mehrere Anbieter zu treffen. Daneben werden mögliche Probleme, die im Zusammenhang mit dem Einsatz einer NLP-Anwendung stehen, betrachtet. Datenqualität oder Datenverfügbarkeit können den Einsatz von NLP-Lösungen erschweren. Außerdem ist auch die Einbindung in bestehende Geschäftsprozesse genau vor dem Hintergrund von zukünftigen Kosten (Integrations- oder Lizenzkosten) und Aufwänden zu bewerten. Schließlich sollte bei der Bewertung der Ergebnisse der NLP-Anwendung die Lokalisierung (Sprache, Region, kulturelle Unterschiede) betrachtet werden.

5. Implementierung der NLP-Services in Geschäftsprozessen

Im finalen Schritt wird die tatsächliche Implementierung der evaluierten und ausgewählten NLP-Services in den Geschäftsprozessen des Unternehmens vollzogen. Dabei bietet sich ein stufenweises Vorgehen (von Prozess zu Prozess) an, um Performance-Gesichtspunkte im Blick zu behalten. Mit der Einführung der Services endet das Vorhaben jedoch nicht, da z. B. die Verfügbarkeit und Qualität der NLP-Services kontinuierlich überwacht werden sollten. Um Lock-in-Effekte minimal zu halten, sollten abstrakte Schnittstellen zu den Anbietern implementiert werden (abstrakte Funktions- und Methodennamen), um diese bei Bedarf einfacher austauschen zu können. Nach der erfolgreichen Implementierung sollten auch die Mitarbeitenden in der Nutzung der NLP-Services und Interpretation der Ergebnisse geschult werden, um die Veränderung fest im Unternehmen zu verankern. Dabei müssen auch Bedenken von Mitarbeitenden proaktiv adressiert werden.

6 Zusammenfassung

Im vorliegenden Beitrag wurden die Möglichkeiten von cloud-basierten KI-Anwendungen aufgezeigt. Diese reichen von Bilderkennung, über Prognosemodelle, über Betrugsprävention, hin zur Verarbeitung von natürlicher Sprache. Nachdem kurz die Grundlagen von NLP dargestellt wurden, wurden typische Anwendungsfälle von sieben NLP-Aufgaben im Kontext von mittelständischen Unternehmen beleuchtet. Um diesen Unternehmen einen Einstieg in die Nutzung von NLP aus der Cloud zu vereinfachen, wurde ein Marktüberblick entwickelt, der final 32 Unternehmen, welche KI-Anwendungen für NLP bereitstellen, präsentiert. Da die Einführung von NLP-Services Unternehmen oftmals vor Herausforderungen stellt, werden diese aufgezeigt und im Rahmen eines fünfstufigen Vorgehens adressiert.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt T. Pröhl, S. Baumgartner und R. Zarnekow geben an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Allahyari M, Pouriye S, Assefi M, Safaei S, Trippe ED, Gutierrez JB, Kochut K (2017) Text Summarization Techniques: A Brief Survey. <https://arxiv.org/pdf/1707.02268>. Zugegriffen: 25.5.2023
- Arnbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M (2010) A view of cloud computing. *Commun ACM* 53:50–58. <https://doi.org/10.1145/1721654.1721672>
- Berente N, Gu B, Recker J, Santhanam R (2021) Managing artificial intelligence. *MISQ* 45:1433–1450. <https://doi.org/10.25300/MISQ/2021/16274>
- Chen, Chiang, Storey (2012) Business Intelligence and Analytics: From Big Data to Big Impact. *MISQ* 36:1165. <https://doi.org/10.2307/41703503>
- Dashtipour K, Poria S, Hussain A, Cambria E, Hawalah AYA, Gelbukh A, Zhou Q (2016) Multilingual sentiment analysis: state of the Art and independent comparison of techniques. *Cogn Comput* 8:757–771. <https://doi.org/10.1007/s12559-016-9415-7>
- Goldberg Y (2022) Neural network methods in natural language processing. Springer, Cham
- Goodfellow I, Courville A, Bengio Y (2016) Deep learning. MIT Press, Cambridge
- Grefenstette G (1999) Tokenization. In: Halteren H (Hrsg) Syntactic wordclass tagging. Springer, Dordrecht, S 117–133
- Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, Zhao L (2019) Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed Tools Appl* 78:15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349:255–260. <https://doi.org/10.1126/science.aaa8415>
- Jurafsky D, Martin JH (2009) Speech and language processing; An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall Pearson Education International, Upper Saddle River
- Khurana D, Koli A, Khatter K, Singh S (2023) Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* 82:3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Lebret RP (2016) Word embeddings for natural language processing. EPFL. <https://infoscience.epfl.ch/record/221430>. Zugegriffen: 25.5.2023
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Li J, Sun A, Han J, Li C (2022) A survey on deep learning for named entity recognition. *IEEE Trans Knowl Data Eng* 34:50–70. <https://doi.org/10.1109/tkde.2020.2981314>
- Manning CD, Schütze H (2005) Foundations of statistical natural language processing. MIT Press, Cambridge
- Neri F, Aliprandi C, Capecci F, Cuadros M, By T (2012) Sentiment analysis on social media. 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
- Onan A, Korukoğlu S, Bulut H (2016) Ensemble of keyword extraction methods and classifiers in text classification. *Expert Syst Appl* 57:232–247. <https://doi.org/10.1016/j.eswa.2016.03.045>
- Rahman A, Ng V (2009) Supervised models for coreference resolution. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 2—EMNLP '09. Association for Computational Linguistics, Morristown

- Rajanak Y, Patil R, Singh YP (2023) Language detection using natural language processing. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE
- Rittinghouse JW, Ransome JF (2009) Cloud computing; Implementation, management, and security. CRC Press, Boca Raton
- Suhaili MS, Salim N, Jambli MN (2021) Service chatbots: a systematic review. *Expert Syst Appl* 184:115461. <https://doi.org/10.1016/j.eswa.2021.115461>
- Velden T, Boyack KW, Gläser J, Koopman R, Scharnhorst A, Wang S (2017) Comparison of topic extraction approaches and their results. *Scientometrics* 111:1169–1221. <https://doi.org/10.1007/s11192-017-2306-1>
- Wei C, Wang Y-C, Wang B, C-CJ K (2023) An overview on language models: recent developments and outlook. <http://arxiv.org/pdf/2303.05759v1>. Zugegriffen: 25.5.2023