# 5 Practical Examples of NLP Use Cases

## Management Summary

In modern companies, information in text form can be found in many places in day-to-day business. Depending on the business context, this can involve invoices, emails, customer input (such as reviews or inquiries), product descriptions, explanations, FAQs, and applications. Until recently, these information sources were reserved mainly for human beings, as the understanding of a text is a technologically challenging problem for machines.

Due to recent achievements in deep learning, several different NLP ("Natural Language Processing") tasks can now be solved with outstanding quality.

In this article, you will learn how NLP applications solve various business problems through five practical examples, which ensured an increase in efficiency and innovation in their field of application.

## Table of Contents

* Introduction
  + What is NLP? – A Short Overview
* The 5 NLP Use Cases
  + Text Classification in the Recruitment Process
  + Usability Improvement of a News Page
  + A Chatbot for a Long-Distance Bus Company
  + A Question-Answering System as a Voice Assistant for Technical Questions About the Automobile
  + Automatic Text Summaries (Text Generation) of Damage Descriptions for a Property Insurance
* Outlook

#h-introduction

#h-what-is-nlp-a-short-overview

#h-the-5-use-cases

#h-text-classification-in-the-recruitment-process

#h-usability-improvement-of-a-news-page

#h-a-chatbot-for-a-long-distance-bus-company

#h-a-question-answering-system-as-a-voice-assistant-for-technical-questions-about-the-automobile

#h-automatic-text-summaries-text-generation-of-damage-descriptions-for-a-property-insurance

#h-outlook

## Introduction

Natural Language Processing (NLP) is undoubtedly an area that has received special attention in the Big Data environment in the recent past. The interest in the topic, as measured by Google, has [more than doubled](https://trends.google.com/trends/explore?date=today%205-y&q=%2Fm%2F05flf) in the last three years. This shows that innovative NLP technologies have long since ceased to be an issue only for big players such as Apple, Google, or Amazon. Instead, a general democratization of the technology can be observed. One of the reasons for this is that according to an IBM estimate, about 80% of "global information" is [not available in structured databases](https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know), but unstructured, natural language. NLP will play a key role in the future when it comes to making this information usable. Thus, the successful use of NLP technologies will become one of the success factors for digitization in companies.

To give you an idea of the possibilities NLP opens up in the business context today, I will present five practical use cases and explain the solutions behind them in the following.

### What is NLP? – A Short Overview

As a research topic that had already occupied linguists and computer scientists in the 1950s, NLP had a barely visible existence on the application side in the 20th century.

The main reason for this was the availability of the necessary training data. Although the availability of unstructured data, in the form of texts, has generally increased exponentially, especially with the rise of the Internet, there was still a lack of suitable data for model training. This can be explained by the fact that the early NLP models mostly had to be trained under supervision (so-called supervised learning). However, supervised learning requires that training data must be provided with a dedicated target variable. This means that, for example, in the case of text classification, the text corpus must be manually annotated by humans before the model training.

This changed at the end of the 2010s when a new model generation of artificial neural networks led to a paradigm shift. These so-called "Language Models" are based on huge text corpora of Facebook, Google, etc., (pre-)trained by randomly masking individual words in the texts and predicting them in the course of training. This is so-called self-supervised learning, which no longer requires a separate target variable. In the course of the training, these models learn a contextual understanding of texts.

The advantage of this approach is that the same model can be readjusted for various downstream tasks (e.g., text classification, sentiment analysis, named entity recognition) with the help of the learned contextual understanding. This process is called transfer learning. In practice, these pre-trained models can be downloaded so that only the fine-tuning for the specific application must be done by additional data. Consequently, high-performance NLP applications can now be developed with little development effort.

To learn more about Language Models (especially the so-called Transformer Models like "BERT", resp. "roBERTa", etc.) as well as trends and obstacles in the field of NLP, please read the article on NLP trends by our colleague Dominique Lade. [https://www.statworx.com/de/blog/neue-trends-im-natural-language-processing-wie-nlp-massentauglich-wird/].

## The 5 Use Cases

**Text Classification in the Recruitment Process**

A medical research institute wants to make its recruitment process of study participants more efficient.

For testing a new drug, different, interdependent requirements are placed on the persons in question (e.g., age, general health status, presence/absence of previous illnesses, medications, genetic dispositions, etc.). Checking all these requirements is very time-consuming. Usually, it takes about one hour per potential study participant to view and assess relevant information. The main reason for this is that the clinical notes contain patient information that exceeds structured data such as laboratory values and medication: Unstructured information in text form can also be found in the medical reports, physician's letters, and discharge reports. Especially the evaluation of the latter data requires a lot of reading time and is therefore very time-consuming. To speed up the process, the research institute is developing a machine learning model that pre-selects promising candidates. The experts then only have to validate the proposed group of people.

### The NLP Solution

From a methodological point of view, this problem is a so-called text classification. Based on a text, a prognosis is created for a previously defined target variable. To train the model, it is necessary - as usual in supervised learning - to annotate the data, in this case the medical documents, with the target variable. Since a classification problem has to be solved here (suitable or unsuitable study participants), the experts manually assess the suitability for the study for some persons in the pool. If a person is suitable, they are marked with a one (=positive case), otherwise with a zero (=negative case). Based on these training examples, the model can now learn the relationships between the persons' medical documents and their suitability.

To cope with the complexity of the problem, a correspondingly complex model called [ClinicalBERT](https://arxiv.org/abs/1904.05342) is used. This is a language model based on BERT (Bidirectional Encoder Representations from Transformers), which was additionally trained on a data set of clinical texts. Thus, ClinicalBERT can generate so-called representations of all medical documentation for each person. In the last step, the neural network of ClinicalBERT is completed by a task-specific component. In this case, it is a binary classification: For each person, a probability of suitability should be output. Through a corresponding linear layer, the high-dimensional text documentation is finally transformed into a single number, the suitability probability. In a gradient procedure, the model now learns the suitability probabilities based on the training examples.

### Further Application Scenarios of Text Classification

Text classification often takes place in the form of sentiment analysis. This involves classifying texts into predefined sentiment categories (e.g., negative/positive). This information is particularly important in the financial world or for social media monitoring. Text classification can also be used in various contexts where it is vital to sort documents according to their type (e.g., invoices, letters, reminders).

## Usability Improvement of a News Page

A publishing house offers its readers on a news page a large number of articles on various topics. In the course of optimization measures, one would like to implement a better recommender system so that for each article, further suitable (complementary or similar) articles are suggested. Also, the search function on the landing page is to be improved so that the customer can quickly find the article he or she is looking for.

To create a good data basis for these purposes, the publisher decided to use Named Entity Recognition (NER) to assign automated tags to the texts, improving both the recommender system and the search function. After successful implementation, significantly more suggested articles are clicked on, and the search function has become much more convenient. As a result, the readers spend substantially more time on the page.

### The NLP Solution

To solve the problem, one must first understand how NER works:

NER is about assigning words or entire phrases to content categories. For example, "Peter" can be identified as a person, "Frankfurt am Main" is a place, and "24.12.2020" is a time specification. There are also much more complicated cases. For this purpose, compare the following pairs of sentences:

1. In the past, Adam didn't know how to parallel park. (park = from the verb "to park")
2. Yesterday I took my dog for a walk in the park. (park = open green area)

It is perfectly evident to humans that the word "park" has a different meaning in each of the two sentences. However, this seemingly simple distinction is anything but trivial for the computer. An entity recognition model could characterize the two sentences as follows:

1. "[**In the past]** *(time reference),* **[Adam]** *(person)* didn’t know how to parallel [park] (verb).”
2. "**[Yesterday]** *(time reference)* **[I]** (person) took my dog for a walk in the **[park]** *(location).*

In the past, rule-based algorithms would have been used to solve the above NER problem, but here too, the machine learning approach is gaining ground:

The present multiclass classification problem of entity determination is again addressed using the BERT model. Additionally, the model is trained on an annotated data set in which the entities are manually identified. The most comprehensive publicly accessible database in the English language is the Groningen Meaning Bank (GMB). After successful training, the model can correctly determine previously unknown words from the context resulting from the sentence. Thus, the model recognizes that prepositions like "in, at, after..." are followed by a location, but more complex contexts are also used to determine the entity.

### Further Application Scenarios of NER:

NER is a classic information retrieval task and is central to many other NER tasks, such as chatbots and question-answer systems. Also, NER is often used for text cataloging, where the type of text is determined based on valid recognized entities.

## A Chatbot for a Long-Distance Bus Company

A long-distance bus company would like to increase its accessibility and expand the communication channels with the customer. In addition to its homepage and app, the company wants to offer a third way to the customer, namely a Whatsapp-Chatbot. The goal is to perform specific actions in the conversation with the chatbot, such as searching, booking, and canceling trips. In addition, the chatbot is intended to create a reliable way of informing passengers about delays.

With the introduction of the chatbot, not only existing passengers can be reached more quickly, but also, contact can be established with new customers\* who have not yet installed an app.

### The NLP solution

Depending on the requirements that are placed on the chatbot, you can choose between different chatbot architectures.

Over the years, four main chatbot paradigms have been tested: In a first generation, the inquiry was examined for well-known patterns and accordingly adapted prefabricated answers were spent ("pattern matching"). More sophisticated is the so-called "grounding", in which information extracted from knowledge libraries (e.g., Wikipedia) is organized in a network by Named Entity Recognition (see above). Such a network has the advantage that not only registered knowledge can be retrieved, but also unregistered knowledge can be inferred by the network structure. In "searching", question-answer pairs from the conversation history (or from previously registered logs) are directly used to find a suitable answer. The use of machine learning models is the most proven approach to generate suitable answers ("generative models") dynamically.

The best way to implement this modern chatbot with clearly definable competencies for the company is to use existing frameworks such as [Google Dialogflow](https://cloud.google.com/dialogflow/?utm_source=google&utm_medium=cpc&utm_campaign=emea-de-all-en-dr-skws-all-all-trial-e-gcp-1009139&utm_content=text-ad-none-any-DEV_c-CRE_461634802071-ADGP_Hybrid%20%7C%20AW%20SEM%20%7C%20SKWS%20~%20EXA_M%3A1_DE_EN_Dialogflow-KWID_43700057038463280-kwd-401718033071-userloc_9044398&utm_term=KW_dialogflow-NET_g-PLAC_&&gclid=Cj0KCQjwxNT8BRD9ARIsAJ8S5xYQrIDEUQ3l0hlfs4DebsM8UfKdX4Gr0KsrJVy9y03sJv4l4NhgjFEaAq8DEALw_wcB). This is a platform for configuring chatbots that have the elements of all previously mentioned chatbot paradigms. For this purpose, parameters such as *intents*, *entities,* and *actions* are passed.

An *intend* ("user intention") is, for example, the timetable information. By giving different example phrases ("How do I get from ... to ... from ... to ...", "When is the next bus from ... to ...") to a language model, the chatbot can assign even unseen input to the correct intend (see text classification).

Furthermore, different travel locations and times are defined as *entities*. If the chatbot now captures an intend with matching entities (see NER), an *action*, in this case a database query, can be triggered. Finally, an intend-answer with the relevant information is given and adapted to all information in the chat history specified by the user ("stateful").

### Further Application Scenarios of Chatbots:

There are many possible applications in customer service, depending on the complexity of the scenario, from the automatic preparation (e.g., sorting) of a customer order to the complete processing of a customer experience.

**A Question-Answering System as a Voice Assistant for Technical Questions About the Automobile**

An automobile manufacturer discovers that many of its customers do not get along well with the manuals that come with the cars. Often, finding the relevant information takes too long, or it is not found at all. Therefore, it was decided to offer a Voice Assistant to provide precise answers to technical questions in addition to the static manual. In the future, drivers will be able to speak comfortably with their center console when they want to service their vehicle or request technical information.

### The NLP solution

Question-answer systems have been around for decades, as they are at the forefront of artificial intelligence. A question-answer system that would always find a correct answer, taking into account all available data, could also be called "General AI". A significant difficulty on the way to General AI is that the area the system needs to know about is unlimited. In contrast, question-answer systems provide good results when the area is delimited, as is the case with the automotive assistant. In general, the more specific the area, the better results can be expected.

For the implementation of the question-answer system, two data types from the manual are used: structured data, such as technical specifications of the components and key figures of the model, and unstructured data, such as instructions for action. All data is transformed into question-answer form in a preparatory step using other NLP techniques (classification, NER). This data is transferred to a version of BERT that has already been pre-trained on a large question-answer data set ("SQuAD"). The model is thus able to answer questions that have already been fed into the system and provide educated guesses for unseen questions.

### Further Application Scenarios of Question-Answer Systems:

With the help of question-answer systems, company-internal search engines can be extended by functionalities. In e-commerce, answers to factual questions can be given automatically based on article descriptions and reviews.

**Automatic Text Summaries (Text Generation) of Damage Descriptions for a Property Insurance**

An insurance company wants to increase the efficiency of its claim settlement department. It has been noticed that some claims complaints from the customer lead to internal conflicts of responsibility. The reason for this is simple: customers usually describe the claims over several pages, and an increased training period is needed to be able to judge whether or not to process the case. Thus, it often happens that a damage description must be read thoroughly to understand that the damage itself does not need to be processed. Now, a system that generates automated summaries is to remedy this situation. As a result of the implementation, the claim handlers can now make responsibility decisions much faster.

### The NLP solution

One can differentiate between two different approaches to the text summary problem: In the *extraction*, the most important sentences are identified from the input text and are then used as a summary in the simplest case. In a*bstraction*, a text is transformed by a model into a newly generated summary text. The second approach is much more complex since paraphrasing, generalization, or the inclusion of further knowledge is possible here. Therefore, this approach has a higher potential to generate meaningful summaries but is also more error-prone. Modern text summary algorithms use the second approach or a combination of both methods.

A so-called sequence-to-sequence model is used to solve the insurance use case, which assigns a word sequence (the damage description) to another word sequence (the summary). This is usually a recurrent neural network (RNN), trained based on text summary pairs. The training process is designed to model the probability of the next word depending on the last words (and additionally, an "inner state" of the model). Similarly, the model effectively writes the summary "from left to right" by successively predicting the next word. An alternative approach is to have the input numerically encoded by the Language Model BERT and to have a GPT decoder autoregressively summarize the text based on this numerical representation. With the help of model parameters, it can be adjusted in both cases how long the summary should be.

### Further Application Scenarios of Text Generation:

Such a scenario is conceivable in many places: Automated report writing, text generation based on retail sales data analysis, electronic medical record summaries, or textual weather forecasts from weather data are possible applications. Text generation is also used in other NLP use cases such as chatbots and Q&A systems.

## Outlook

These five application examples of text classification, chatbots, question-answer systems, NER, and text summaries show that there are many processes in all kinds of companies that can be optimized with NLP solutions.

NLP is not only an exciting field of research but also a technology whose applicability in the business environment is continually growing.

In the future, NLP will not only become a foundation of a data-driven corporate culture but also already holds a considerable innovation potential through direct application, in which it is worth investing.

At STATWORX, we already have years of experience in the development of customized NLP solutions. Here are two of our case studies on NLP: [Social Media Recruiting with NLP](https://www.statworx.com/de/data-science/data-science-case-studies/social-media-recruiting-nlp/) & [Supplier Recommendation Tool](https://www.statworx.com/de/data-science/data-science-case-studies/supplier-recommendation-tool/). We are happy to provide you with individual advice on this and many other topics.