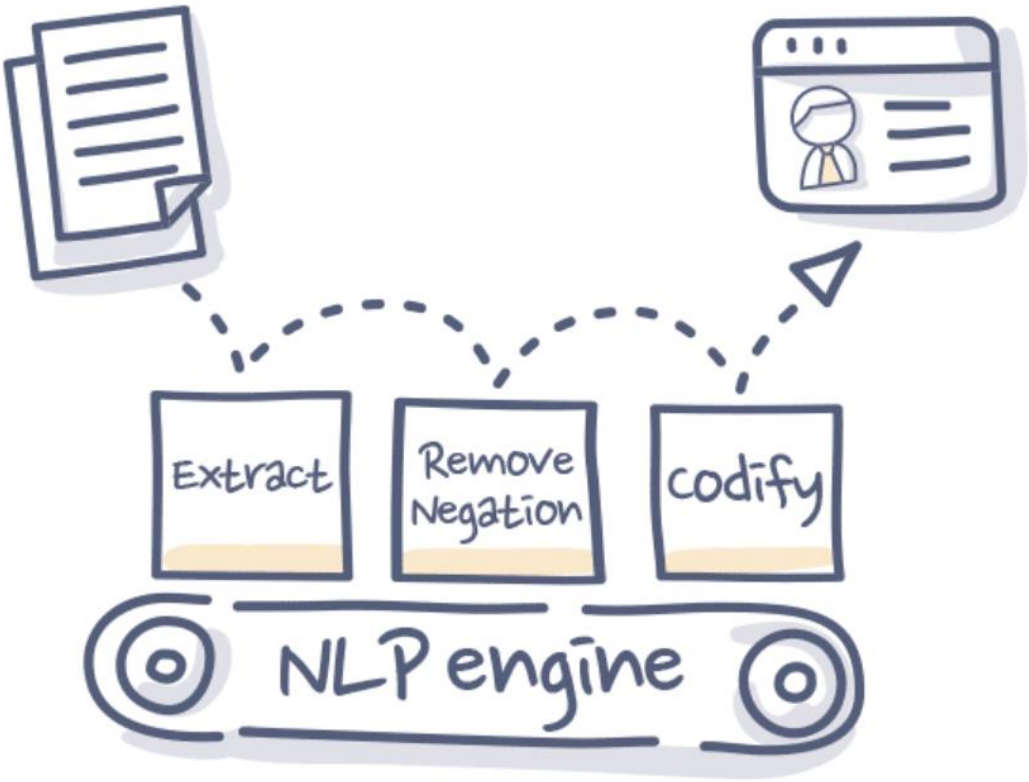
Natural Language Processing (NLP) system for detecting insights in unstructured biomedical using Azure AI

Problem Statement

We implement a **Natural Language Processing (NLP) system for detecting insights in unstructured biomedical.**

Natural Language Processing (NLP): This is a subfield of **AI that focuses on enabling computers to understand, interpret, and generate human language**.

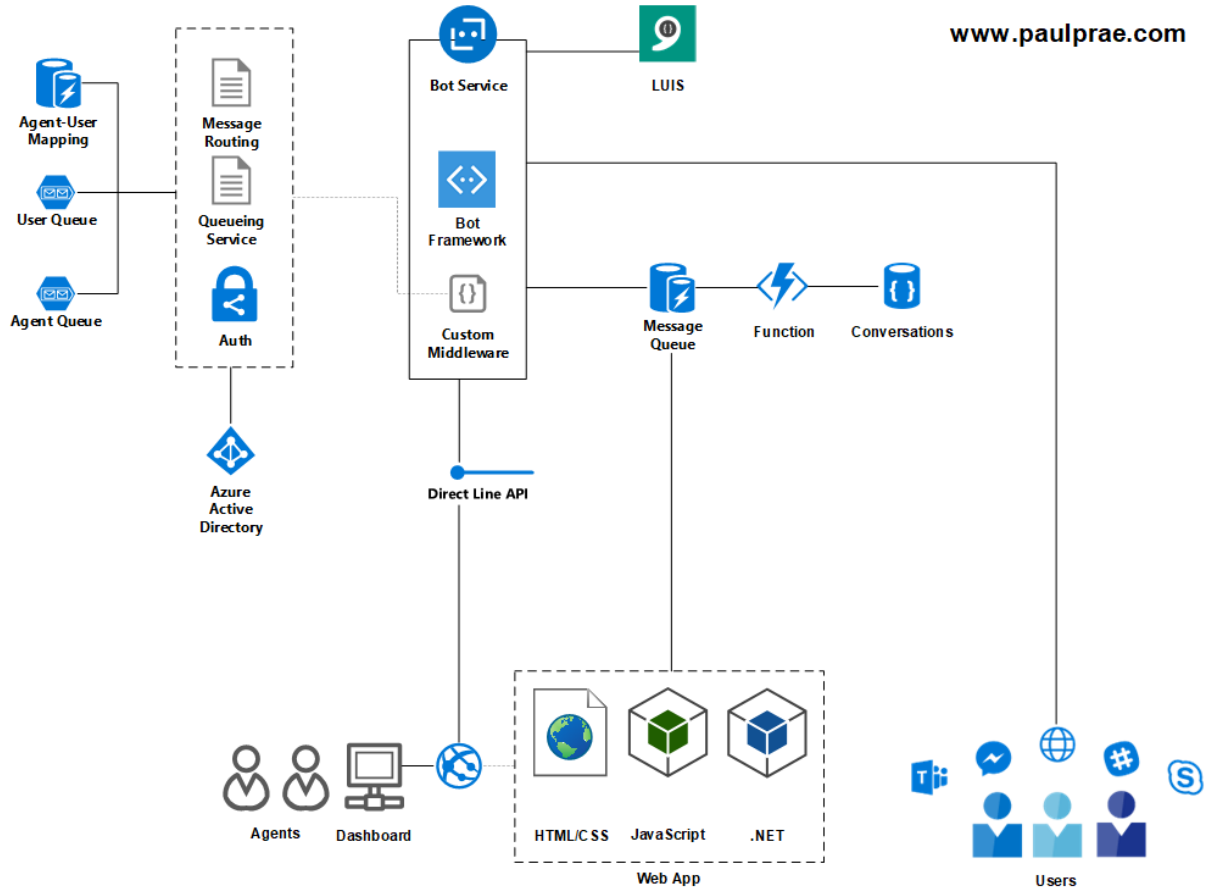
NLP can be used to extract information from unstructured biomedical texts, such as research papers, clinical notes, and electronic health records.

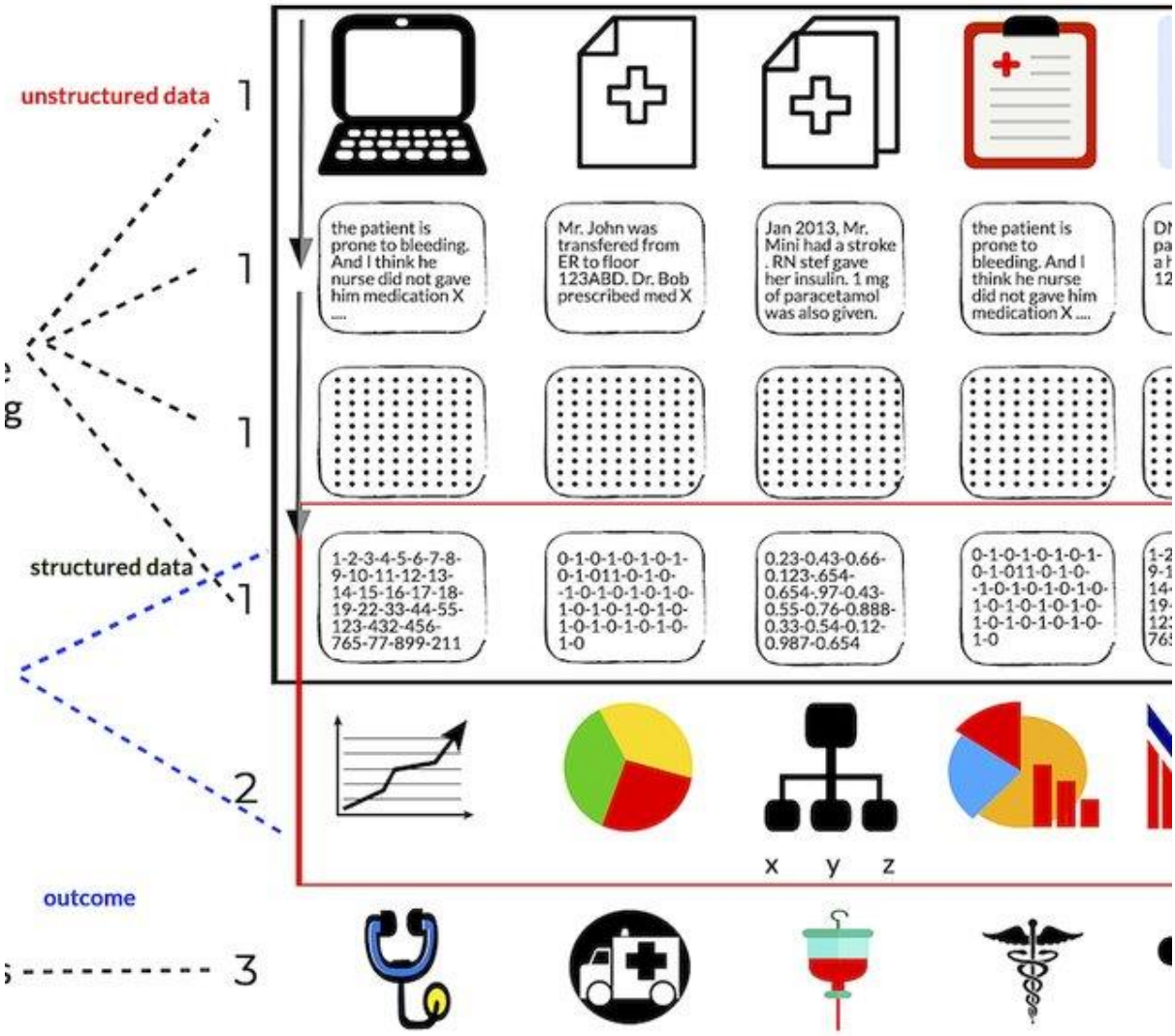


Solution/Architecture

To build a natural language processing (NLP) system for detecting insights in unstructured biomedical data, you can follow these steps:

1. **Collect and pre-process the text data:** First, you will need to collect the unstructured biomedical data that you want to analyse, and pre-process it to prepare it for NLP. This may involve cleaning the data, removing any irrelevant or redundant information, and formatting it in a way that is suitable for NLP.
2. **Use NLP tools and techniques to extract insights from the text:** There are many NLP tools and techniques that you can use to extract insights from unstructured biomedical data, such as tokenization, part-of-speech tagging, and named entity recognition. You can use these tools and techniques to identify key phrases, sentiments, and language patterns in the text data.
3. **Use machine learning to build and train a model:** Once you have extracted insights from the text data using NLP, you can use machine learning to build and train a model to classify the text data or make predictions based on it.
4. **Deploy the model:** Once you have trained your model, you can deploy it to a production environment so that it can be used to analyse and make predictions on new unstructured biomedical data.
5. **Monitor and maintain the system:** Finally, you will need to monitor the performance of your NLP system and make any necessary updates or adjustments to ensure that it is functioning correctly.


www.paulprae.com
Agent-User Mapping
User Queue
Agent Queue
Message Routing
Queueing Service
Auth
Azure Active Directory
Bot Service
LUIS
Bot Framework
Custom Middleware
Direct Line API
Message Queue
Function
Conversations
Agents
Dashboard
HTML/CSS
JavaScript
.NET
Web App
Users



To decode natural language processing (NLP), you can use various techniques and tools that are designed to analyse and interpret human language. Some common NLP techniques include:

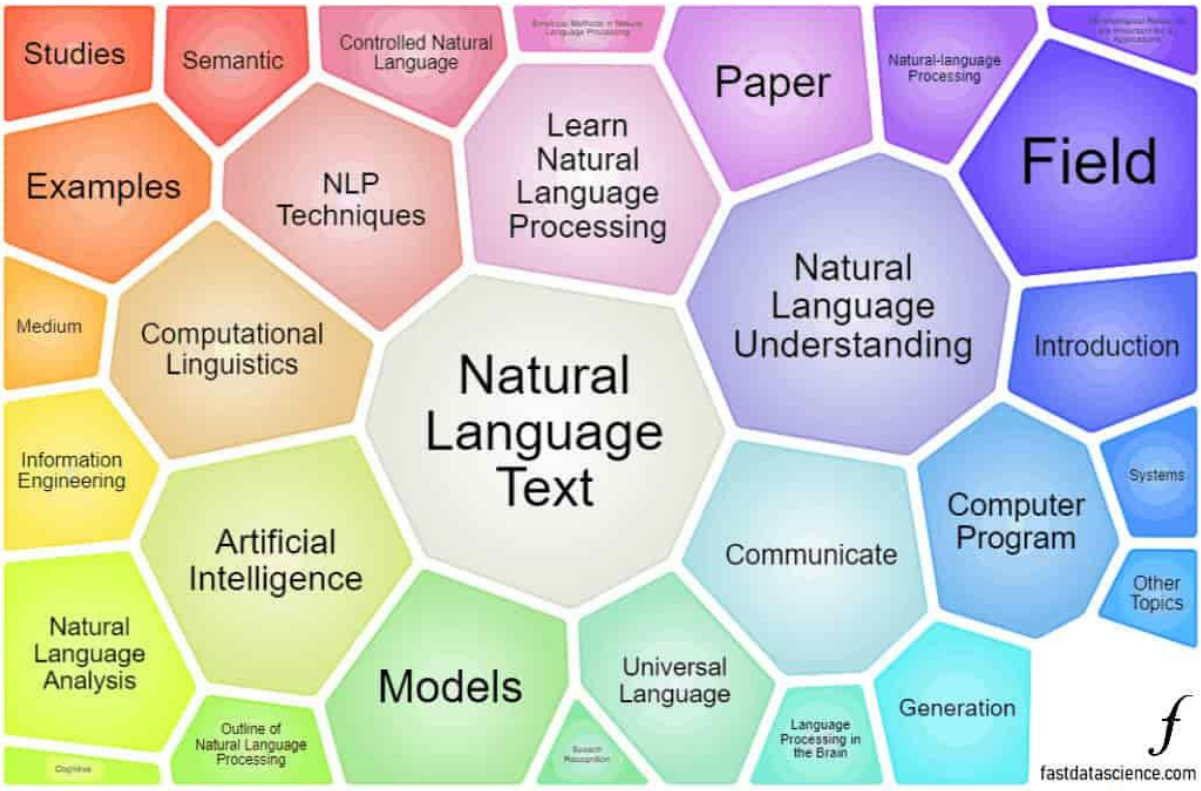
1. **Tokenization:** This is the process of breaking down a piece of text into smaller units called tokens, which can be words, phrases, or symbols. Tokenization is a key step in many NLP tasks, as it allows the text to be analysed at a more granular level.

2. **Part-of-speech tagging:** This is the process of identifying the part of speech (e.g., noun, verb, adjective) of each token in a piece of text. Part-of-speech tagging is useful for understanding the grammatical structure of a sentence and can be used to help disambiguate words with multiple meanings.

3. **Named entity recognition:** This is the process of identifying and extracting named entities (e.g., people, organizations, locations) from a piece of text. Named entity recognition can be used to extract structured information from unstructured text.

4. **Stemming:** This is the process of reducing a word to its base form, called a stem. Stemming is useful for grouping together words that have a common meaning, even if they are spelled differently.

5. **Syntactic parsing:** This is the process of analysing the syntactic structure of a sentence and determining the relationships between the words and phrases in it. Syntactic parsing can be used to understand the meaning of a sentence and extract information from it.



NLP processes unstructured data from different sources (e.g., EMRs, literature, and social media) so that analytics systems can interpret it (Figure 1). Once NLP converts the text to structured data, health systems can use it to classify patients, extract insights, and summarize information.

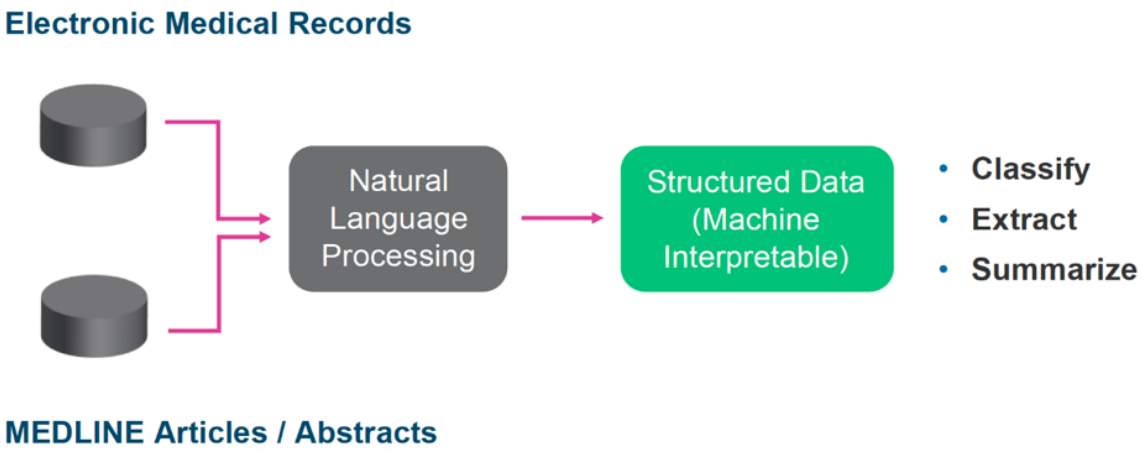


Figure 1: NLP classifies, extracts, and summarizes unstructured text

Four areas in which healthcare NLP can improve function—and, ultimately, care—include EHR usability, predictive analytics, phenotyping, and quality improvement:

NLP Improves EHR Data Usability

The typical EHR arranges information by patient encounter, making it difficult to find critical patient information (e.g., social history—a strong predictor of readmissions). NLP can enable an EHR interface that makes patient encounter information easier for clinicians to find.

By organizing the interface into sections, and including words associated with concerns patients described during encounters, the interface populates the rest of the page with information related to that word. For example, all mentions of fatigue would show on a timeline at the top of the page, and the notes about the word would show in a box at the bottom of the page. The interface makes it easier for clinicians to find buried data and make diagnoses they might have otherwise missed.

NLP Enables Predictive Analytics

One of the more exciting benefits of NLP is its ability to enable predictive analytics to improve significant population health concerns. For example, according to recent reports, **suicide** has been rising the United States. Healthcare professionals are working to understand who is at risk so they can intervene. A 2018 **study** used NLP to predict suicide attempts by monitoring social media. Results showed clear indicators of imminent suicide attempts by Twitter users who posted fewer emojis in text, limited emojis to certain types (e.g., blue or broken heart symbols), or increased postings of angry or sad tweets prior to attempting suicide (Figure 2). The system had a 70 percent prediction rate with only a 10 percent false positive rate.

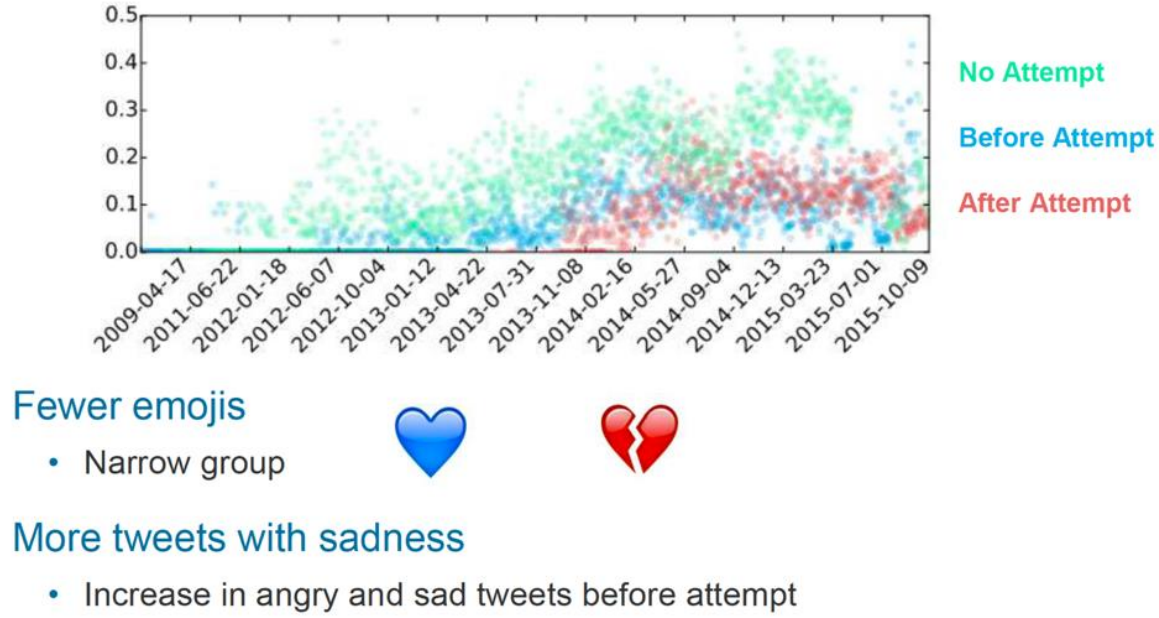


Figure 2: Using NLP to recognize suicide risk in emoji use

NLP Boosts Phenotyping Capabilities

Phenotype is an observable physical or biochemical expression of a specific trait in an organism. These traits may be related to appearance, biochemical processes, or behaviour. Phenotyping helps clinicians' group or categorize patients to provide a deeper, more focused look into data (e.g., listing patients who share certain traits) and the ability to compare patient cohorts. Currently, most analysts and clinicians use structured data for phenotyping because it's easy to extract for analysis. NLP gives analysts a tool to extract and analyse unstructured data (e.g., follow-up appointments, vitals, charges, orders, encounters, and symptoms), which some experts estimate makes up 80 percent of all patient data available. Access to unstructured data makes a lot more information available to create phenotypes for patient groups.

NLP also allows for richer phenotypes. For example, pathology reports contain a lot of information, such as a patient's condition, location of a growth, stage of a cancer, procedure(s), medications, and genetic status. While traditional analytics can't access that data from pathology reports, NLP empowers analysts to extract this type of data to answer complex, specific questions (e.g., cancerous tissue types associated with certain genetic mutations).

NLP Enables Health System Quality Improvement

The federal government and associated agencies require all hospitals to report certain outcome measures. One required measure is adenoma detection rate (ADR), which is the rate at which doctors find adenomas during a colonoscopy. The current process for reporting is to pay someone to analyse a small sampling of patient charts, read through the pathology reports, and calculate the ADR. NLP automates and accelerates this process, increasing the sample size of patient charts and allowing real-time analysis.

A clinician has developed a **report card** that uses NLP to automatically calculate ADR. Studies show that when physicians can see quantifiable results of their performance, they tend to change their behaviour. In this case, physicians who receive feedback about their ADR changed their behaviours to improve detection rate. This is important because for every 1 percent increase in ADR, there is a 3 percent decrease in colon cancer mortality.

While the four areas in which NLP enhances the value of healthcare data show significant promise, NLP has a long way to go to widespread adoption and a large-scale impact on outcomes improvement.

Technical Details and Implementation of solution

To implement a natural language processing (NLP) system using Azure AI for detecting insights in unstructured biomedical data, you can follow these steps:

1. **Identify the problem and define the objectives:** The first step in implementing an NLP system using Azure AI is to identify the problem that the system will be used to solve and define the objectives that the system should achieve.
2. **Collect and pre-process the data:** The next step is to collect and pre-process the unstructured biomedical data that will be analysed by the NLP system. This may involve cleaning the data, removing any irrelevant or redundant information, and formatting it in a way that is suitable for the Azure AI algorithms.
3. **Use Azure Text Analytics to extract insights from the text:** Once the data has been pre-processed, you can use Azure Text Analytics to extract insights from the text. Azure Text Analytics is a cloud-based service that uses NLP to identify key phrases, sentiments, and language patterns in text data.
4. **Use Azure Machine Learning to build and train a machine learning model:** Once you have extracted insights from the text data using Azure Text Analytics, you can use Azure Machine Learning to build and train a machine learning model to classify the text data or make predictions based on it.
5. **Evaluate the performance of the model:** Once the model has been trained, you can evaluate its performance and select the best model for deployment.
6. **Deploy the model:** The final step is to deploy the selected model to Azure, where it can be used to analyse and make predictions on new unstructured biomedical data. Ongoing maintenance and monitoring will also be required to ensure that the model continues to function correctly.

Training the NLP model for different languages

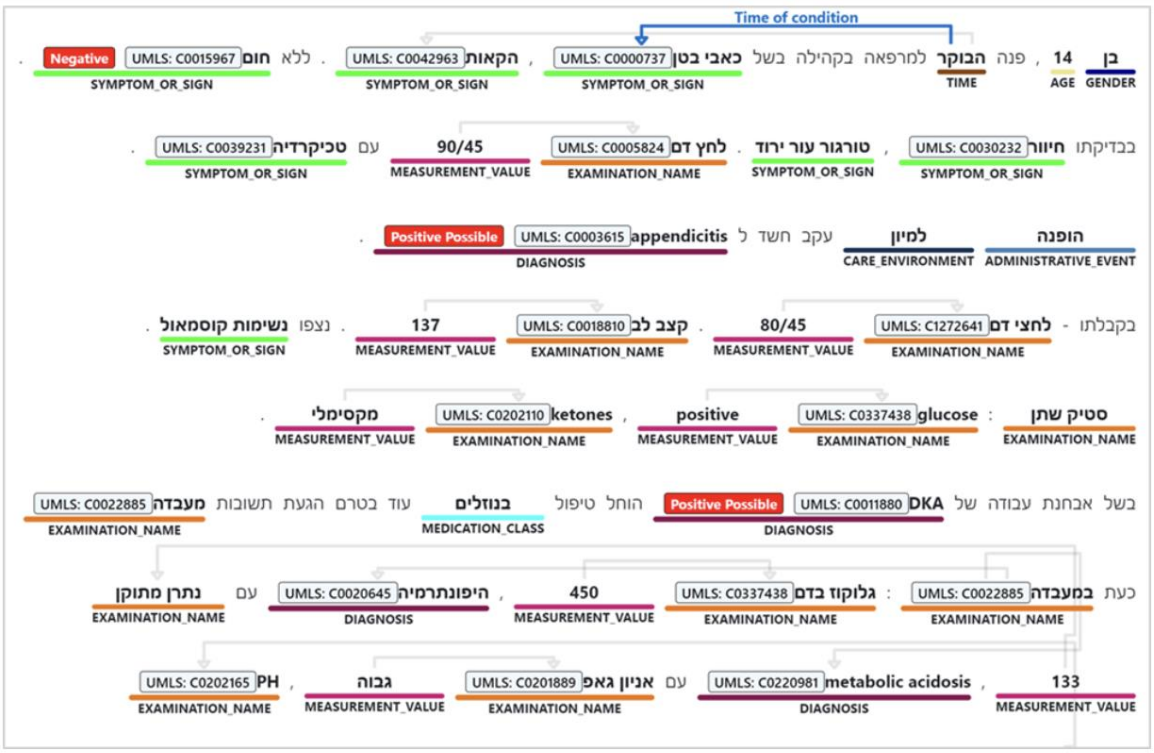


Figure 1: Analysis of Hebrew unstructured biomedical text using Text Analytics for Health

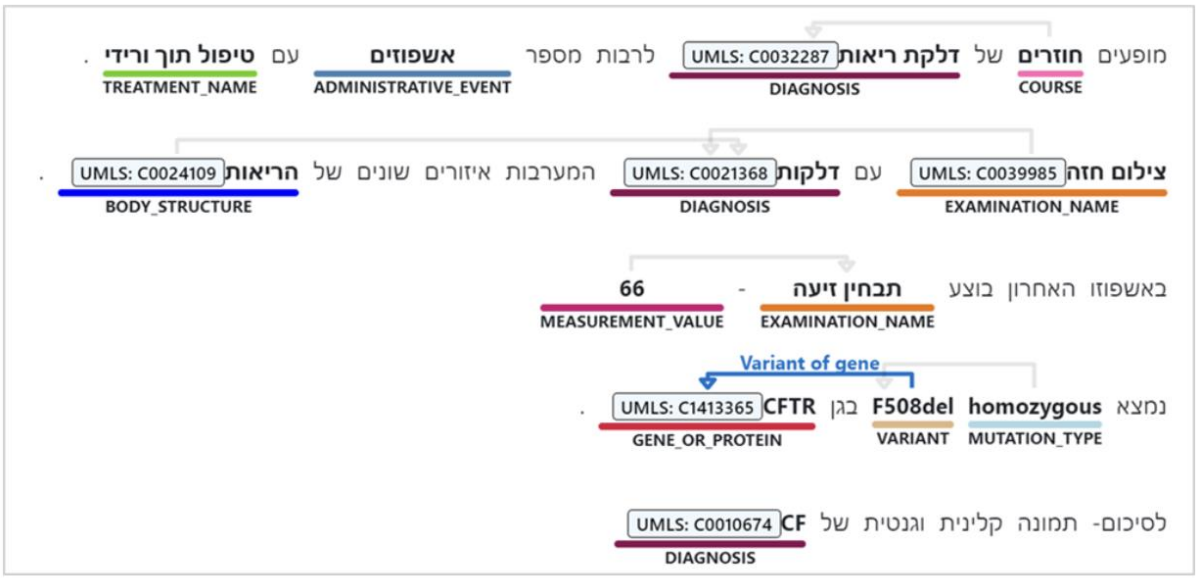


Figure 2: Analysis of Hebrew unstructured biomedical text using Text Analytics for Health

Analysing unstructured text for Real-World Data

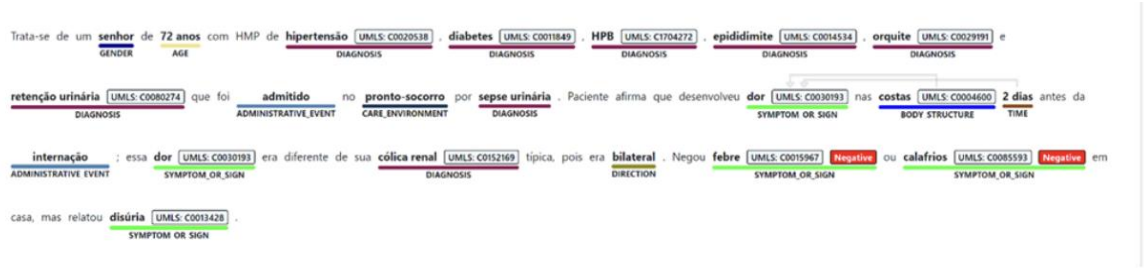


Figure 3: Analysis of Portuguese unstructured biomedical text using Text Analytics for Health

Analysis and structuring to Fast Healthcare Interoperability Resources (FHIR®)

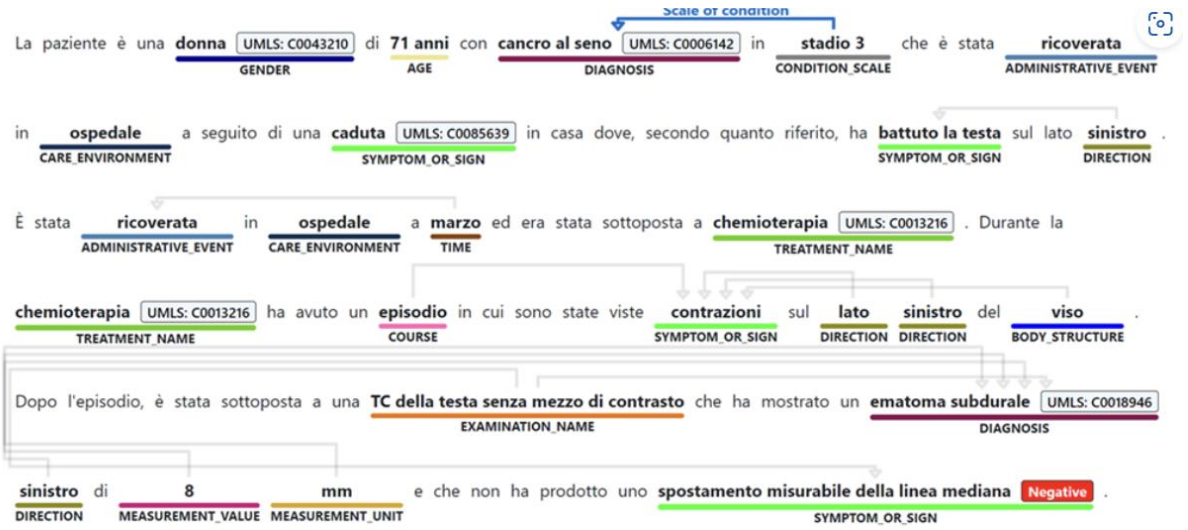
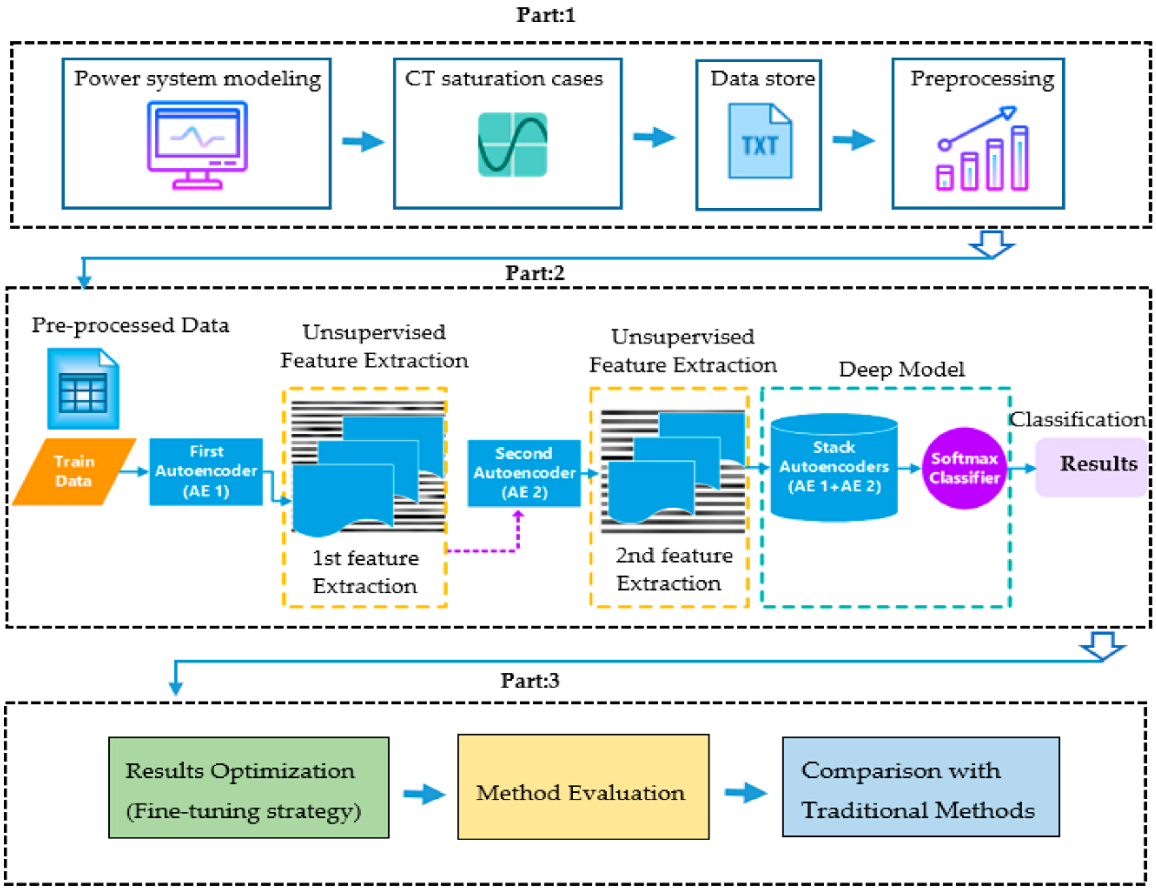


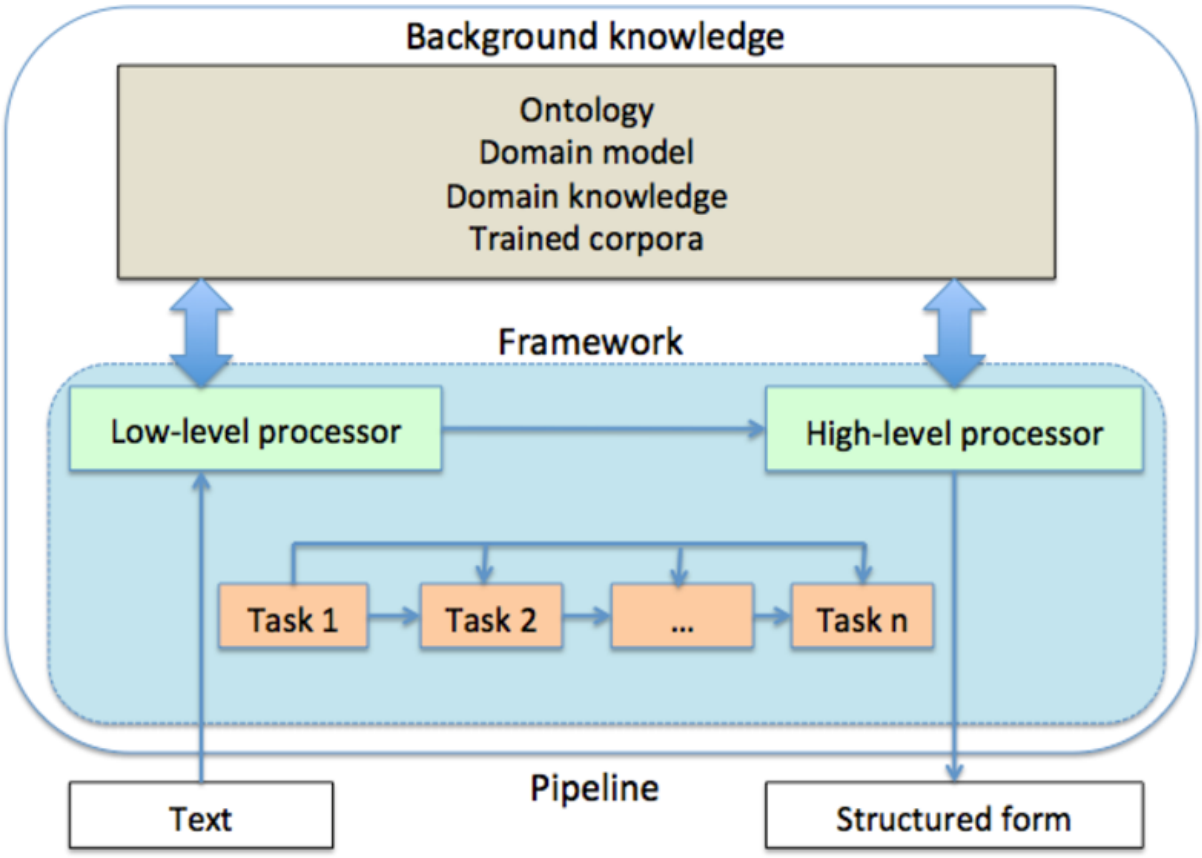
Figure 4: Analysis of Italian unstructured biomedical text using Text Analytics for Health

Challenges in implementing the solution

There may be several challenges involved in implementing a natural language processing (NLP) system using Azure AI for detecting insights in unstructured biomedical data, including:

1. **Data quality and quantity:** One of the main challenges in implementing an NLP system is ensuring that the data is of high quality and enough. Poor quality data, or data that is not representative of the problem being solved, can lead to poor performance of the NLP system.
2. **Pre-processing the data:** Pre-processing the data to prepare it for NLP can be a time-consuming and labour-intensive process, especially if the data is large or poorly formatted.
3. **Feature extraction:** Extracting relevant features from the data using NLP techniques can be a challenging task, especially if the data is noisy or contains a large amount of irrelevant information.
4. **Model training:** Training machine learning models on the extracted features can be a complex and time-consuming process and may require significant computing resources.
5. **Model deployment:** Deploying the trained model to a production environment, such as Azure, can be a complex process that requires careful planning and testing.
6. **Model maintenance**: Ongoing maintenance and monitoring of the deployed model will be required to ensure that it continues to function correctly and provide accurate insights.


Part:1
Power system modeling
CT saturation cases
Data store
TXT
Preprocessing
Part:2
Pre-processed Data
Train Data
First Autoencoder (AE 1)
Unsupervised Feature Extraction
1st feature Extraction
Second Autoencoder (AE 2)
Unsupervised Feature Extraction
2nd feature Extraction
Deep Model
Stack Autoencoders (AE 1+AE 2)
Softmax Classifier
Classification
Results
Part:3
Results Optimization (Fine-tuning strategy)
Method Evaluation
Comparison with Traditional Methods


Background knowledge
Ontology
Domain model
Domain knowledge
Trained corpora
Framework
Low-level processor
High-level processor
Task 1
Task 2
...
Task n
Pipeline
Text
Structured form

Business Benefit

A natural language processing (NLP) system using Azure AI can provide several business benefits, including:

1. **Improved efficiency and productivity:** An NLP system can automate tasks that would otherwise be done manually, such as data entry or analysis, freeing up time for more value-added activities.
2. **Enhanced decision-making:** An NLP system can analyse large amounts of data and provide insights and recommendations that can help inform decision-making.
3. **Improved customer experience:** An NLP system can be used to analyse customer feedback and identify trends or patterns that can help improve the customer experience.
4. **Increased revenue:** An NLP system can help businesses identify new opportunities, such as new markets or product offerings, that can drive revenue growth.
5. **Cost savings:** An NLP system can help businesses reduce costs by automating tasks, improving efficiency, and reducing errors.
6. **Competitive advantage:** An NLP system can give businesses a competitive edge by enabling them to make faster, more informed decisions and respond more quickly to changes in the market.