



Case study

Leveraging NLP in Medicine with Custom Model for Scientific Reports Analysis

Our client is an international medical research organization that helps manufacturing companies gather insights necessary for delivering safe and effective medical products

To enhance their research capabilities and speed up data analysis, they wanted to integrate innovative artificial intelligence (AI) functionality into their existing data analysis processes.

The challenge

The client asked us to create an efficient AI tool that could process documents in PDF and XML formats and provide a concise analysis of the data they contain in the form of a PDF report.

That was the start of a long-term collaboration that has resulted in many improvements to the platform's feature set, stability, and user experience.

The result

Apriorit provided the client with an efficient AI-powered tool that automates the processing of scientific reports. The natural language processing (NLP) model we developed can:

- Analyze reports provided in PDF and XML formats
- Extract valuable and unobvious insights from processed data
- Summarize findings in a comprehensive format that is easy to work with

With this solution in place, the client now gets to leverage the benefits of using NLP in medicine.

Our approach

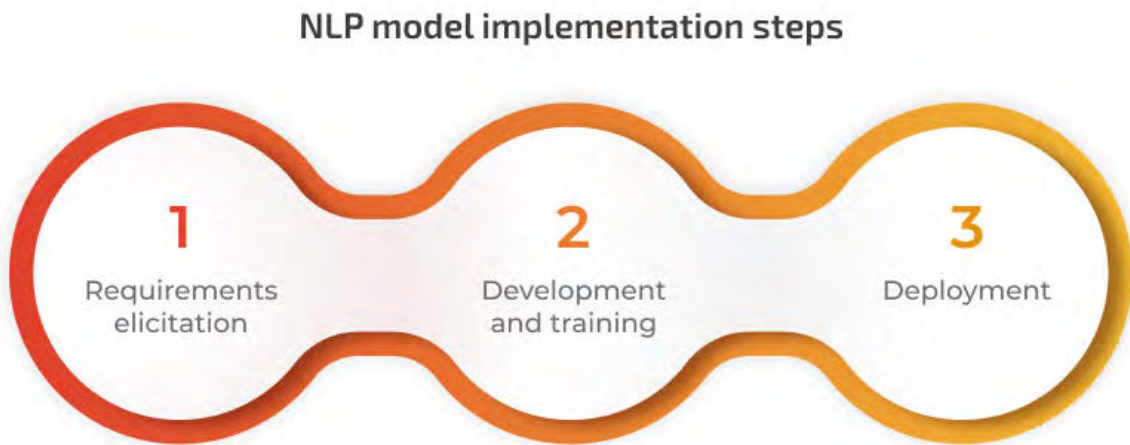
For this project, Apriorit formed a dedicated team that researched possible implementation scenarios and, once approved by the client, designed an NLP model for document processing and analysis.

Our dedicated team consisted of AI developers along with a data engineer, quality assurance specialist, business analyst, and project manager. Core technologies the team worked with while building the NLP model included GPT-4, spaCy, text embedding algorithms, and document parsing tools.

| NLP model development project details | |
|---|--|
| Team | Technology stack |
| <ul style="list-style-type: none">• AI developers• Data engineer• Quality assurance specialist• Business analyst• Project manager | <ul style="list-style-type: none">• GPT-4• spaCy• Text embedding algorithms• Document parsing tools |

How we did it

The Apriorit AI engineering team created an NLP model for automating and enhancing the processing of scientific research reports in three steps:



1. Requirements elicitation

Having a clear, unified vision with finalized product requirements is mandatory for selecting the most fitting development approach, tools, and technologies. To better understand client expectations and finalize requirements for the final product, we involved an experienced business analyst.

Our specialist started by investigating similar applications of NLP in the medical field and working closely with project stakeholders, to outline their vision and expectations. Then, we elicited and finalized project requirements and planned further steps of the development process.

At this point, we determined that it would be more cost-efficient to build the solution using a pre-trained NLP model and moved to the implementation stage.

2. Development and training of an NLP AI model

There are different options for implementing natural language processing in medicine. Working with pre-trained AI models is an efficient way to speed up the development process and optimize the budget without sacrificing a model's accuracy and flexibility.

As a basis for this project, we used GPT-4 and trained the model on the client's data. To ensure high accuracy of the customized model, we pre-processed the data and composed large training and testing datasets out of reports provided by the client.

As the client's analysts work with scientific reports provided in PDF and XML formats, we only used these formats in our datasets.

During the training and fine-tuning stages of NLP model development, we also worked closely

3. Solution deployment

After fine-tuning the model, the Apriorit team assisted the client with deploying the solution. Documentation provided by our specialists also helped the client educate their analysts on proper and efficient use of the developed AI tool.

The impact

Apriorit specialists built, fine-tuned, and integrated an NLP-powered solution into the client's system. The model employed can analyze and summarize scientific reports and medical research with over 92.8% accuracy.

Extracting medical information from clinical text with NLP enabled the client to accelerate scientific research while allowing their analysts to focus on more complex tasks.