

## **Transform extensive documentation to a user-friendly question-answering system by leveraging advanced Natural Language Processing techniques and latest AI technology.**

### **Proposal:**

We will provide your company with the opportunity to unlock the full potential of your extensive documentation through our cutting-edge service. This service transforms complex texts into a user-friendly question-answering system. By leveraging advanced natural language processing techniques and the latest AI technology, we simplify, streamline, and make accessing information more intuitive. There's no need to rely on ChatGPT or any external services outside your firewall. Our system is designed to be deployed and run within your own SaaS environment, ensuring complete control and security. With our solution, you'll face no licensing restrictions, security concerns, or limitations on the number of documents processed, types of documents, or the quantity of tokens that feed the model.

### **Our Goal:**

To ensure that customers from large and small corporations, non-profit organizations, legal firms, news agencies, engineers, researchers, product developers, sales and marketing teams will find joy in engaging with documentation, rather than investing considerable time navigating through a myriad of unstructured documents and manuals.

### **Some background:**

The advent of large language models (LLM) like GPT (Generative Pre-trained Transformer) series, represents a significant leap forward in artificial intelligence (AI), opening up new possibilities for human-computer interaction and understanding. As the technology continues to evolve, it presents both opportunities and challenges that will shape the future of AI and its role in society. These models have dramatically transformed how machines understand, generate, and interact with human language, leading to advancements across a wide range of applications.

LLMs can generate concise summaries of extensive documents, making it easier to grasp the main points without reading the entire text. This is particularly useful for engineers, researchers, legal professionals, and anyone needing to quickly understand the essence of lengthy reports, manuals, user guides, articles, or contracts.

LLMs can be used to develop question-answering (Q/A) systems that allow users to ask natural language questions about the content of a large document. A Q/A system democratizes access to information by allowing users with varying levels of expertise to query complex documents using natural language. This means that one doesn't need to be an expert in a specific domain or skilled at keyword-based search techniques to find the information they need.

### **Why we make a difference:**

A widely known set of models is provided by OpenAI, those that are used in ChatGPT application. However, while OpenAI models offer powerful capabilities for enhancing

applications with AI, it's essential to address security, privacy, and compliance concerns. In many cases processed documents cannot be exported outside the company's firewall. Running all processes inside a local Software as a Service (SaaS) system presents a number of benefits for organizations prioritizing data control, security. The approach we provide is engaged with:

- Enhanced Data Privacy and Security
- Customization and Control
- Regulatory Compliance
- Reliable Performance and Availability
- Cost Predictability
- Simplified IT Landscape
- With your developers we will create a custom Q/A System: Tailored solutions that integrate seamlessly with your existing data repositories with all types of documents, providing real-time, accurate answers to complex queries.
- We can convert documentations, user guides and manuals of your company products into user-friendly Q/A system.
- Tuning up your system depending on your needs and customization.
- Help your developers create, maintain and improve your inhouse system, which will run on your internal SaaS environment.
- No licensing issues. Models used are under Apache 2.0 license.
- Your software and operation engineers will have full control on developing, bug fixing and maintaining this service.
- To establish a customer workflow, the system requires minimal resources and can leverage GPU clusters provided by AWS and similar cloud services. Additionally, it's compatible with private cloud environments.
- For development, debugging and bug fixing, the system is designed to operate on standard CPU-based laptops equipped with 32-64GB of RAM.

At a high level, the system's workflow includes the following steps:

1. The system picks documents from specified sources, such as files or web documents or pages.

2. A database is created that stores these documents. There might be different types of DBs. The easiest and straightforward way is to utilize a Vector-DB. In this case no special arrangements required. It is created in a special form of DB index and stored once.
3. Setup a transformer model and connect it with the DB.
4. In your tenant start user-friendly conversation.
5. Every tenant will run its own question-answering conversation against the system running with the backend docker container. You can cancel conversation any time and start a new one. A running conversation caches conversational context modified with every query. This context is essential to remember a flow of a discussed subject.

Additionally, an important benefit to note is that the Q/A system serves as a valuable tool for document writers, aiding in the enhancement and refinement of the documents they create.

### **Our team**

**Vitaly Bulgakov:** <https://nanoandgiga.com/board/bulgakov.html>

AI Development and Application, Computer Science, Software Engineering, Data Science.

**Vladislav Gorbich:** <https://nanoandgiga.com/board/gorbich.html>

Web software development, information Technology.

**Anatoli Korkin:** <https://nanoandgiga.com/board/korkin.html>

Project management and customer relations