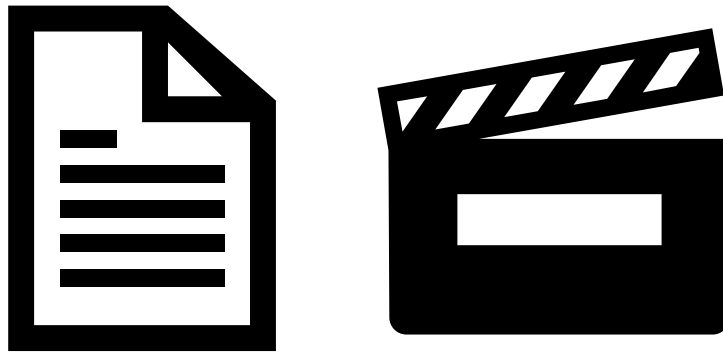# Business Proposal

Text-to-Video software for Zenya

By

Podrono, Wensley
Dumitru, Andreea
Krasnodębski, Andrzej
Harbi, Mohammed
Al-Eryani, Mohammed

Document version history

Version	Date	Changes
0.1	15/02/2023	Initial text
0.2	08/02/2023	Points 1-10, final version
0.3		

Document distribution history

Person	Date	Notes

Table of Contents

1. Management Summary

This business proposal presents a solution idea to the problem reported by the Infoland company. Zenya, which is a core software platform in the clients offer, seeks improvement in the area of change introduction and management to support employee adaptation to this change and continous improvement. Currently, those processes are done by sending an email with plain instructions which according to Infoland may and should be optimized to make the process smoother for both employers and employees (both Zenya clients). The solution would be implemented into Zenyas' Boost module which is responsible for monitoring change activities.

Green AI company comes with a possible solution to the reported problem and after careful investigation proposes the following idea:

Introducing an Artificial Intelligence solution together with a pipeline which would analyze a given input text (an email with specified change) and render a user-friendly informative video instead of plain text email. Meeting the clients requirements, the aforementioned short video clip would be a smooth transfer of details in a easy to memorize manner. Tracking the behaviour afterwars would be a task for the Boost module and is out of scope of this project.

This documment describes in detail the given problem, research and work planning, together with a broader specifications for the solution itself.

2. Introduction

In today's fast-paced work environment, people often find it difficult to dedicate a lot of time to learn and understand more complex concepts. This is where short informative videos turn out to be incredibly helpful and convenient to consume.

Our client – Infoland, aims to implement this strategy in one of their core software products - Zenya. The idea in mind, is to create an AI solution which would automatically convert text emails into user-friendly, engaging and informative videos. The purpose is to help Zenya's user adapt to a change in their workplace.

Green AI - a team of Fontys University students will carefully analyze the given problem and requirements. Afterwards, come up with a possible solution and plan to make this idea happen. Our uttermost goal is to satisfy Infoland's requirements by solving their issue and contribute to Zenya's improvement.

Thank you for considering this proposal and input to the solution.

3. Clients' Company Information

Founded in 1998 as a small smart software startup, Infoland has grown to a team of over 100 specialists. Together they design and develop software solutions for businesses in a variety of markets. Their solutions focus on safeguarding themes such as quality, safety, the environment and the well-being of people. Achieving this by providing professionals with the tools to quickly find the right information, perform risk analyses, document and learn from mistakes, and more. They believe the key to successful software is a happy end user and aim to achieve this by paying close attention to what the professionals actually want and providing them with reliable, easy to use solutions.

Zenya is the largest software suite offered by Infoland. It encompasses tools to support and optimize the quality-and risk management of organizations. Zenya enables easy document management, quick incident reporting, thorough auditing processes, and a complete overview of risk management within an organization. Despite the high customer satisfaction rating, they always look for ways to improve. By extending functionality, improving the user experience, optimizing performance, and looking for ways to innovate and keep Zenya up to date on modern technologies.

4. Problem Understanding

a) Presented Problem

At the moment there is a new module being developed within Zenya. This module is called Boost. The purpose of this module is to boost the level of knowledge for users, influencing a change in behaviour, in a way that is as much as nonintrusive as possible. One of the use cases is that an important change in a protocol or instruction gets known by users, understood by users, and is brought to practice by users. It is not sufficient to only inform users of such a change. Another use case is that there isn't a specific change but for some other reason there is an initiative to focus on a certain topic, for example as safety theme or the month of cybersecurity. This may be the case because of an (near) incident or just because that safety theme is an annual recurring theme. Boost lets you use a set of tools that you can use to bring small learning nuggets/micro learnings to users spread over time. For some users information is easier to comprehend through a video than in plain text. That's why we want to explore the possibilities to generate a video with spoken text that is based on a piece of text. This can be an avatar, AI generated person or AI deep fake based on an actual person.

In this project the goal is to develop working concept that provides a video based on a Dutch piece of text.

b) Presented Requirements

- Research suitable techniques and set up one or more Proof of Concept (POC)
- Integration with Zenya is not part of this project
- Cost management, control costs using existing AI/NLP services when used on Zenya SaaS infrastructure by Infoland customers
- The solution should be scalable to support the current 1.5M users of Zenyaand future growth The solution is required to support multiple tenants (organizations)
- The POC is not required to support multiple languages at first but it should be suitable to support multiple languages in the future. For this POC it should support English or Dutch, preferably Dutch.
- The solutions must be GDPR compliant
- The solution must be deployable on Azure

5. Goal of the Project

The goal of this project is to help Infoland in achieving their target Zenya improvement by introducing a TEXT-TO-VIDEO Artificial Intelligence solution. With this project, our client aims to provide a valuable resource to learners of all ages and backgrounds, helping them to quickly grasp and retain knowledge on a wide range of topics.

We can accomplish it by creating a text-to-video AI model together with a process pipeline that will generate short, educational and easy to consume videos for Zenya software. The model should be able to take written text script and convert it into clips of maximum length of 1 minute. The videos should be concise, visually appealing, and should effectively communicate complex concepts in a short amount of time.

A more detailed description of the desired video concept/style and theme will be available together with a solution breakdown and explanation in later stage of this project. It concerns a more technical issues and approach while the purpose of this document is to present a business proposal of the initial findings and the general idea.

6. Research

This section describes the research part of Green AI project. It contains research questions linked to a methodology in which it should be persued and planning for those activities. Please note that a separate research document will be available with the results of our findings.

a) Main Research Question

How can a text-to-video AI model be developed to fit this use case?

b) Sub Research Questions

SQ.1 How does the desired solution look like?

SQ.2 How does the input text look like?

SQ.3 What are the specifics of the video?

SQ.4 Who is the end-user of the video?

SQ.5 How to break down the desired solution into smaller pieces?

SQ.6 How to automate the designed pipeline?

SQ.7 What AI models are neccesary to develop the end product?

SQ.8 What kind and amount of data is required to train those models individually?

SQ.9 How to obtain this specified data?

SQ.10 Which required models are already existing and can be re-used and which are still needed to be created?

SQ.11 How to evaluate the performance of the pipeline and video?

c) Research Methodology

Green AI uses different research methodologies from the ICT Research Methods and plans to apply them as following:

Expert Interview: **SQ.1, SQ.2, SQ.3, SQ.4, SQ.11**

Brainstorm: **SQ.1, SQ.3, SQ.5, SQ.6, SQ.7, SQ.11**

Decomposition: **SQ.5**

Available product analysis: **SQ.7, SQ.10, SQ.9**

Literature study: **SQ.6, SQ.7, SQ.8, SQ.11**

d) Research Planning

Please find the deadlines for reach research question below. This plannign allows us to carefully investigate vital issues and proceed with the project.

Research Question	Deadline
SQ.1, SQ.2, SQ.4	27/02/2023
SQ.3, SQ.5	08/03/2023
SQ.6, SQ.7	13/03/2023
SQ.8, SQ.9, SQ.10	15/03/2023
SQ.11	20/03/2023

7. Possible Solutions

a) Brainstorming Results

There are several models that can be used to approach the problem of generating videos from text input, including sequence-to-sequence models, generative adversarial networks (GANs), and transformer models. Each model has its advantages and disadvantages, and the choice will depend on the specific needs and goals of the pipeline.

Sequence-to-sequence models are a type of neural network that can map an input sequence to an output sequence. They have been used for various natural language processing tasks, including text-to-speech and machine translation. One approach to using sequence-to-sequence models for text-to-video generation is to map the textual input to a sequence of video frames. However, this approach can be challenging, as it requires handling multiple modalities, such as text, image, and video.

Generative adversarial networks (GANs) are another type of neural network that have shown promise in generating realistic images and videos. GANs consist of two neural networks, a generator and a discriminator. The generator creates fake data, while the discriminator tries to distinguish the fake data from real data. The training process is a competition between the generator and discriminator, with the goal of producing high-quality fake data that can fool the discriminator. One potential approach to using GANs for text-to-video generation is to train the generator to create video frames based on textual input, with the discriminator providing feedback on the quality of the generated frames.

Transformer models are a type of neural network architecture that has been shown to be highly effective for natural language processing tasks. Transformer models are based on a self-attention mechanism that allows the model to focus on relevant parts of the input sequence when making predictions. One approach to using transformer models for text-to-video generation is to map the textual input to a sequence of image or video features, and then use the transformer model to generate a sequence of video frames based on the features.

Each of these models has its strengths and weaknesses, and the choice will depend on the specific needs and goals of the project. Sequence-to-sequence models have been widely used in natural language processing tasks and can be effective for generating sequences of data but may struggle with the complexity of the text-to-video generation problem. GANs can generate highly

realistic images and videos but may require large amounts of data and computational resources to train. Transformer models are highly effective for natural language processing tasks but may require additional steps, such as mapping the textual input to image or video features, to be effective for text-to-video generation.

A more detailed description of the possible video appearance and theme will be available together with a solution breakdown and explanation in later stage of this project. It concerns a more technical issues and approach while the purpose of this document is to present a business proposal of the initial findings and the general idea.

b) EDA Results

As no data have been provided up front from the client, Green AI will have to collect it in the first phase of this project. Exploratory Data Analysys (EDA) follows data collection and will be docummented in a separate file.

8. Scope

a) Features IN scope

The scope of this project includes the development of a text-to-video model that can take written transcripts and convert them into informative and engaging videos. Working proof of concept is a generalized way to describe the scope of this project.

The model will be trained on a large dataset of text and video pairs, and the videos will be generated using a pre-existing dataset or graphics and animations.

The model will be evaluated using a set of metrics to measure its performance, and optimizations will be made to improve the quality of the generated videos. The spoken voice is English.

The final model will be tested on a new test dataset to ensure its quality and will be deployed in a production environment for real-world use on Zenya.

b) Features OUT OF scope

This project will not include integration with Zenya software as well as implementing language support other than English.

c) Deliverables

This section provides an overview of the project's deliverables, including their expected outcomes and delivery timelines. Its purpose is to assist the student team in achieving successful project planning and execution while managing client expectations to ensure their satisfaction with the project's outcomes.

1. **Project Proposal**

A document initiating the project, describing its client, problem, research, approach, and solution. Its aimed to give a detailed yet overview of the situation and planning. Vital to align on the solution and kick off the development.

2. Technical Documentation

Comprehensive documentation that describes the design, implementation, and usage of the product. The technical documentation includes an overview of the product, its architecture, data used, ML (Machine Learning) models used, API, and any other relevant technical details. Delivered alongside the final product delivery. The technical documentation may be updated and revised throughout the product's lifecycle as needed.

Technical documentation includes the following sections released in this order:

- EDA Report
- Data Preparation Report
- Modeling Report

3. Final Report

A comprehensive report that summarizes the results of the project and supplies recommendations for future improvements. The report includes an overview of the project goals, a summary of the project tasks and deliverables, a description of any challenges or obstacles that were met during the project, an analysis of the project results and outcomes, and recommendations for future improvements.

The report makes ample use of visual information, such as charts, graphs, and diagrams, to support the project findings and analysis.

The report analyzes the project from an ethical and legal perspective, using resources such as tict.io to guide the analysis. Delivered at the end of the project, after all deliverables have been completed and the final product has been delivered. It has yet to be determined when the final project report will be delivered.

Final Report is delivered together with training materials that help users understand how to use the product effectively.

4. Final product

The definitive version of the product that has been developed and evaluated according to the project requirements. A fully functional and tested product that meets the client's needs and requirements. Delivered at the end of the project, after all development, testing, and quality assurance work has been completed.

d) Phase Planning

The project timeline for the development of the text-to-video model is 18 weeks, with the following project phases:

Phase 1 –Initiation, Research & Planning

First phase to start the project, gather information and other requirements from the client and start researching given problem. Establishing all facts and clear image creation for approach and goal. Followed by stating the goal and agreeing on the desired output, finished by planning of required activities.

Phase 2 – Data Sourcing & Exploration

Collecting and experimenting with required data. A vital step highly influencing next phases and final outcome.

Phase 3 – Development & Modeling

Main development phase where ideas from the previous ones becomes real and the first versions of the product come to light.

Phase 4 – Evaluation & Optimization

Measuring the performance of current achievements and optimizing a product in following iterations.

Phase 5 – Final Testing & Delivery

Final phase of this project where the best performing model is finally tested and delivered to the client. Transfer of knowledge and documents also happens here. After all steps are finalized, the project can conclude.

9. Communication with the client

Communication with a client is crucial in any project or business relationship, and it plays a vital role in the success of the project. Effective communication ensures that the client's needs and expectations are clearly understood and addressed, and it helps to build a strong and positive relationship between the client and the service provider.

Clear and timely communication with the client helps to establish trust and transparency. It allows the client to have a clear understanding of the project's progress, any challenges that may arise, and how they will be addressed. It also provides an opportunity for the client to provide feedback and input, which can help to improve the final product.

Additionally, communication helps to ensure that the project stays on track and within budget. Regular updates and progress reports allow both the client and the service provider to track progress and make any necessary adjustments to ensure that the project is delivered on time and within the agreed-upon budget.

Communication will be based on a regular update meeting and information exchange via email.

10. Conclusion

The proposed text-to-video model will be a key component in Zenya's AI-powered strategy to provide high-quality informative content to a wide yet specified audience. The model will convert written explanations of complex concepts into engaging and informative videos easily understandable by learners of all ages and backgrounds. With the development of this model, Zenya aims to provide a valuable resource to people, helping them to quickly grasp and retain knowledge on a wide range of topics.

In conclusion, I would like to thank you for taking the time to consider our business proposal. We are confident that our proposed solution will meet your needs and help you achieve your business goals. Our team is dedicated to delivering high-quality results and providing exceptional customer service and support throughout the project. We are excited about the opportunity to work with you and help take your business to the next level.

Please do not hesitate to contact us if you have any questions or if you would like to schedule a consultation. Thank you again for considering our proposal.

We look forward to the opportunity to work with you and help your business evolve,
Team Green AI